Four Essays on the Econometric Analysis of High-Frequency Order Data

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum politicarum
(Doktor der Wirtschaftswissenschaft)

eingereicht an der
Wirtschaftswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
M. Sc Ruihong Huang
geboren am 27.08.1976 in Fujian, China

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Wirtschaftswissenschaftlichen Fakultät:
Prof. Oliver Günther, Ph.D.

Gutachter:

1. Prof. Dr. Nikolaus Hautsch
2. Prof. Dr. Christian T. Brownlees

Tag des Kolloquiums: 14 Februar, 2012

Abstract

Electronic limit order markets account for a large and increasing percentage of global financial trading. Understanding the complex interaction between traders' limit order submission strategies and the state of the market, and their consequent effects on market quality becomes increasingly important for investors, as well as to those who regulate and design automated markets.

In four essays, this thesis examines the aforementioned interaction by econometric analysis of the market impact of limit order, the properties of order flow and traders' (hidden) order submission decisions.

Chapter 1 looks at the market impact of limit orders. Quantifying short-term and long-term effects of limit order submissions on quotes in Euronext, we show that limit orders have significant information content, and how strongly limit orders signal the market depends on both their characteritics (price and size) and the state of limit order books (LOBs).

Chapter 2 provides new empirical evidence on order submission activities and market impacts of limit orders at NASDAQ. We find that traders dominantly submit small size limit orders and cancell most of them immediately after submission. Based on the estimated market impact of orders, we propose a method to predict the optimal size of a limit order conditional on its position in the LOB and a given fixed level of expected impact.

Chapter 3 analyzes traders' decisions on using undisclosed orders in opaque markets. Employing Totalview message data at NASDAQ, we show that market conditions affect traders' order submission strategies and thus the location of hidden liqudity is predictable given observable market characteristics. Out evidence also suggests that traders balance their hidden order placements to compete for the provision of liquidity and protect themselves against picking-off risk.

Chapter 4 presents a program framework for reconstructing LOBs as well as extracting order flow information from message stream data. We design the basic modules of the system in an abstract layer based on common order events in limit order markets, so that it can be easily adapted to data at any limit order markets. The underlying data structure is highly optimized and the programs in these modules are exhaustively tested to guarantee the high reliablity and efficiency of the system.

Zusammenfassung

Transaktionen auf vollständig elektronischen, *Limit-Order*-getriebenen Märkten machen einen großen und wachsenden Teil aller weltweiten Finanzmarkttransaktionen aus. Es ist deshalb von herausragender Bedeutung für Investoren und Regulatoren dieser Märkte, die komplexen Interaktionen zwischen dem Zustand des Marktes und den Limit-Ordern der Marktteilnehmer zu verstehen.

Das Kapitel 1 befasst sich mit den Auswirkungen von Limit-Ordern der Marktteilnehmer auf den Zustand des Marktes und die Marktqualität. In der Analyse quantifizieren wir die kurz- und langfristigen Effekte der Limit-Order Platzierung auf Preisquotierungen. Am Beispiel des Börsenplatzes Euronext zeigen wir, dass eine Limit-Order signifikante Informationen für Preisquotierungen enthält und illustrieren inwieweit der Markt von den Charakteristika der Limit-Order (Preis und Volumen) und dem Zustand des Orderbuchs abhängt.

Das Kapitel 2 enthält neue empirische Resultate über die Limit-Order Aktivität und den Markteinfluss von Limit-Ordern an der New Yorker NASDAQ Börse. Wir dokumentieren, dass Marktteilnehmer hauptsächlich die Platzierung von Limit-Ordern mit kleinen Volumina präferieren, diese aber häufig sofort nach ihrem Einsatz wieder löschen. Basierend auf der geschätzten Marktauswirkung einer individuellen Limit-Order schlagen wir eine Methode zur Prognose des optimalen Volumens einer Limit-Order vor. Die optimalen Eigenschaften der Limit-Order hängen dabei von der Position im Orderbuch sowie der vorab spezifizierten erwarteten bzw. präferierten Marktauswirkung ab.

Im Kapitel 3 werden die Limit-Order-Strategien von Marktteilnehmern in intransparenten Märkten untersucht. Unter Benutzung des Totalview message Datensatzes des NASDAQ Börsenplatzes zeigen wir, dass die Position der sogenannten versteckten Liquidität im Orderbuch von diversen Variablen abhängt, die den Zustand des Marktes beschreiben. Insbesondere zeigen wir, dass die Position der versteckten Liquidität prognostizierbar ist. Die Daten suggerieren zudem, dass Händler die Platzierung sogenannter Hidden-Orders im Hinblick auf günstige Liquidität am Markt und dem "'Picking-Off"'-Risiko ausbalancieren.

Im letzten Kapitel 4 präsentieren wir ein Softwaresystem zur Rekonstruktion von Orderbüchern und zur Extrahierung von Orderflussinformationen aus message stream Daten für Limit-Orders. Die Basismodule des Systems beruhen auf allgemeinen Orderbuch-Ereignissen. Sie sind abstrakt gehalten und können so einfach auf beliebige Märkte mit elektronischen Orderbüchern angewendet werden. Die grundlegende Struktur ist im Hinblick auf Anwendbarkeit optimiert und gründlich getestet worden um eine hohe Verlässlichkeit und Effizienz zu gewährleisten.

Acknowledgment

I owe a measure of gratitude to several people, without whom this thesis would not have been completed.

First of all, I am very grateful to my supervisor, Nikolaus Hautsch, for his inspiring and encouraging advice. I very much benefited from numerous insightful discussions with him. He offered an excellent and exciting research environment at his Chair.

I am greatly indebted to my second advisor, Christian Brownlees for his valuable input. The discussion with him was not only academically fruitful but also personally enjoyable.

This thesis would not have been possible without the many insights that I learned from my colleagues and friends working in the financial industry. I especially appreciate the enlightened talk, the strong expression of interest and the valuable constructive criticism that they have given me. I wish to acknowledge the following persons by name: Andrew Ferraris (UBS Investment Bank), Lada Kyj (Barclays Capital), Daniel Nehren (JP Morgan Chase), Artem Novikov (Trading Physics), Roel Oomen (Deutsche Bank) and Dmitry Shakin (GSA Capital).

Many thanks are due to my colleagues at the "Lehrstuhl für Ökonometrie" who actively commented on my research: Bernd Droge, Axel Groß-Klußmann, Gustav Haitz, Franziska Lottmann, Peter Malec, Tomas Polak, Julia Schaumburg and Melanie Schienle. I also owe special debt to the entire LOBSTER development team including Gustav, Tomas and two student assistants, Jonas Haase and Gagandeep Singh.

I heavily benefited from the financial support by Deutsche Bank and from the stimulating atmosphere within the Quantitative Products Laboratory (QPL). I especially appreciate the impressive scientific activities organized by two scientific directors, Ulrich Horst and Peter Bank, and the operation official, Almut Birsner. Besides, I owe debt to all colleagues in QPL for giving me the wonderful experience both in my academic career and in my social life. Among them, I would like to especially thank Gökhan Cebiroglu, Mstislav Elagin, Antje Fruth, Peter Kratz, Felix Naujokat, Michael Paulsen and Nick Westray for all inspiring discussions during working time and lunchs.

Finally and most of all, I would like to thank my wife, Xi, who not only consistently supported and encouraged me during the last four years but also brought me my sweet daughter, Sophie. They gave me the mental support, optimism, and retreat that enabled me to successfully finish my dissertation.

Contents

List of Figures

List of Tables

Introduction

Electronic limit order markets, which collect traders' orders and automatically match them on the basis of specified priority rules, become increasingly popular in financial markets around the world. Nowadays most equity and derivative exchanges are either pure electronic limit order markets, e.g. NYSE Arca, BATS, Euronext, Australian Stock Exchange (ASX) and Direct Edge, or at least allow for customer limit orders in addition to on-exchange market making, e.g. NASDAQ, NYSE and the London Stock Exchange (LSE). Consequently, understanding the mutual effects between electronic limit order market design features and traders' order submission strategies, and their consequent impact on market quality, are becoming increasingly important to investors, exchanges and regulators.

A *limit order* is an instrument to trade up to a given amount of a security at the best price available, but no worse than the specified *limit price*. Hence, it is an ex ante pre-commitment made by the submitter and is in force until the order is completely filled or cancelled. Limit orders are executed when traders on the other side of the market submit market orders or marketable limit orders. In particular, a *market order* is an instruction to trade a given amount of a security at the best price currently available in the market and a *marketable limit order* is a limit order with such an aggressive limit price that it can be immediately (possibly partially) executed when the trader submits it. Unlike Walrasian markets which apply a uniform market-clearing price to all periodical aggregated orders, limit order markets execute orders discriminatorily, i.e. each limit order executed in a transaction (by a market order) is filled at its respective limit price.

Limit order markets maintain the open limit orders by using a central limit order book (LOB). It typically matches traders' orders on a price and time priority basis. Price priority means that the limit orders offering better limit prices, i.e. limit buys at higher prices and limit sells at lower prices, are executed before limit orders at worse prices. Time priority means that, at each price, older limit orders are executed ahead of more recent limit orders. Based on these order precedence rules, market order traders can trade directly with limit orders supplied by other traders. This direct interaction implies a public supply of liquidity and distinguishes limit order book markets from dynamic dealer markets in which liquidity is only supplied by registered market makers administering each transaction.

Apart from *non-execution risk*, i.e. prices moving away from their order after the submission, and *adverse selection (picking-off) risk*, i.e. prices moving against their newly established position after the execution, large buy-side traders in general face an additional *exposure risk*. When they show their trading intention in the market, *defensive* traders may refrain from trading with them and *parasitic* traders may exploit the option value of large limit orders by front-running them. Consequently, the non-execution risk increases. Moreover, due to new incoming information, previously submitted limit orders may become mis-priced. Fast traders may pick off these orders quickly before submitters can cancel them. This results in a significant increase of picking-off risk.

In order to encourage large traders to actively supply the liquidity in markets, many electronic stock exchanges choose to reduce the pre-trade transparency by allowing traders to hide a proportion of their order sizes. Correspondingly, they typically impose a secondary order precedence rule: exposed orders or exposed parts of undisclosed orders gain time priority over the hidden part of undisclosed orders.

In this thesis, we look at empirical evidence of the complex interaction between traders' limit order submission strategies and the state of the LOB, as well as the underlying economic reasoning. In particular, employing high-frequency order data, we conduct econometric analysis of the market impact of limit orders, characteristics of order flow, and traders' undisclosed order submission decisions conditional on observable market conditions.

Chapter 1 is joint work with my supervisor, Nikolaus Hautsch, and is published in the *Journal of Economic Dynamics & Control*. In this chapter, we quantify the short-run and long-run price effect of posting a limit order in a limit order market by proposing a high-frequency cointegrated VAR model for quotes and order book depths. Estimating impulse response functions based on data from 30 stocks traded at Euronext Amsterdam we show that limit orders have significant market impacts. The strength and direction of quote responses depend on the incoming orders' aggressiveness, their sizes and the state of the book. The effects are qualitatively stable across the market. Cross-sectional variations in the magnitudes of price impacts are well explained by the underlying trading frequency and relative tick size.

In chapter 2, we provide new empirical evidence on order submission activity and price impacts of limit orders at NASDAQ. Employing NASDAQ TotalView-ITCH data, we find that market participants dominantly submit limit orders with sizes equal to a round lot. Most limit orders are cancelled almost immediately after submission if not getting executed. Moreover, only very few market orders walk through the book, i.e. directly move the best ask or bid quote. Estimates of impulse-response functions on the basis of a cointegrated VAR model for quotes and market depths allow us to quantify the market impact of incoming limit orders. We propose a method to predict the optimal size of a limit order conditional on its position in the book and a given fixed level of expected market impact. This chapter is joint work with Nikolaus Hautsch and is published on the conference proceeding of "Market Microstructure, Confronting Many Viewpoints".

Trading under limited pre-trade transparency becomes increasingly popular on financial markets. In Chapter 3, we provide first evidence on traders' use of (completely) undisclosed orders in electronic trading. Employing TotalView-ITCH data on order messages at NASDAQ, we propose a simple method to conduct statistical inference on the location of hidden depth given the state of the market. We show that market conditions reflected by the bid-ask spread, (visible) depth, recent price movements and trading signals affect traders' decisions where to post hidden orders. Our evidence suggests that traders optimize their hidden order placements to (i) compete for the provision of (hidden) liquidity and (ii) protect themselves against adverse selection, front-running as well as "hidden order detection strategies" used by high-frequency traders. Overall, our results show that hidden liquidity is predictable given observable market characteristics and is a key element in modern trading and execution strategies. This chapter is joint work with Nikolaus Hautsch.

The rise of algorithmic trading in electronic limit order markets creates considerable challenges for researchers, who have to cope with extremely large amounts of trading data produced daily by exchanges. In chapter 4, we present a program framework for reconstructing LOBs as well as extracting order flow information from message stream data. The system is modularized based on common order events in the generalized order-processing of limit order markets, so that it can be easily adapted to data from any limit order markets. Moreover, the underlying data structures in the basic modules are highly optimized and algorithms are exhaustively tested to guarantee the reliability of the output data and the efficiency of the entire system. This chapter is joint work with my colleague Tomas Polak, and a software treating the NASDAQ TotalView-ITCH data is implemented by the LOBSTER development team and is accessible for all researchers via `http://lobster.wiwi.hu-berlin.de`.

Chapter 1

Market Impact of a Limit Order

This chapter is based on Hautsch and Huang (2012b).

1.1 Introduction

It is well known that the revelation of trading intention adversely affects asset prices. As also confirmed by theoretical studies[1], passive order placement through limit orders incurs significant market impact even if the order is not been executed. In financial practice, the risk to 'scare' and to ultimately shift the market by limit order placements is well-known and is taken into account in trading strategies. As a consequence, liquidity provision through hidden orders, which allow traders to partly (or entirely) conceal order volume, has gained popularity. However, despite the importance of limit order strategies in modern trading, the actual impact of an incoming (visible) limit order on the subsequent price process is still hardly explored and quantified. In fact, while the analysis of the price impact resulting from a trade is a classical topic in traditional market microstructure research (see, e.g., Dufour and Engle, 2000; Engle and Patton, 2004; Hasbrouck, 1991), empirical evidence on the market impact of limit order placements is addressed by only few recent studies as Eisler, Bouchaud, and Kockelkoren (2011) and Cont, Kukanov, and Stoikov (2011).

This chapter aims at filling this gap in the literature and addresses the following empirical research questions: (i) How strong is the short-run and long-run impact of an incoming limit order in dependence of its position in the book, its size and the state of the book? (ii) Are ask and bid quote responses to incoming limit orders widely symmetric or is there evidence for an asymmetric re-balancing of the book? (iii) How different is the market impact of a limit order compared to that caused by a trade of similar size? (iv) How stable are these effects across the market and do they depend on stock-specific characteristics, such as the underlying trading intensity, minimum tick size and average trade size?

[1] See, e.g., Parlour and Seppi (2008), Boulatov and George (2008) and Rosu (2010).

We propose modelling the processes of ask and bid quotes as well as several levels of depth volume on both sides of the market in terms of a cointegrated vector autoregressive (VAR) model. This framework allows us to study the price impact of limit orders by means of impulse response functions. Each limit order is represented by a shock disturbing the multivariate system of quotes and depths and influencing it dynamically over time. Designing the shock vectors in a specific way allows us to characterize the type of the limit order represented by its size and its position in the order queue as well as the current state of the book.

The motivation for using a cointegrating system stems from the fact that ask and bid quotes are naturally integrated and tend to move in locksteps. Cointegration analysis reveals a stationary linear combination of bid and ask quotes which closely resembles the bid-ask spread. The idea of jointly modelling ask and bid quote dynamics in terms of a cointegrated system originates from Engle and Patton (2004) based on the work of Hasbrouck (1991) and has been used in other approaches, such as Hansen and Lunde (2006) and Escribano and Pascual (2006). Our setting extends and modifies this approach in two major directions: Firstly, we model quotes and depth simultaneously. This yields a novel type of order book model capturing not only quote and depth dynamics but implicitly also dynamics of midquotes, midquote returns, spreads, spread changes as well as order book imbalances. Secondly, we model the system not only on a trade-to-trade basis but exploit the complete order arrival process. Therefore, the model captures all relevant trading characteristics in a limit order book market and thus provides a complete description of the order book in a range close to the best quotes. Hence, the model is particularly useful for liquid assets where most of the market activity is concentrated at the best quote levels. In this sense, the approach complements dynamic models for order book curves such as proposed by Härdle, Hautsch, and Mihoci (2009) and Russell and Kim (2010).

The proposed quote and depth model is estimated by Johansen's (1991) full information maximum likelihood estimator using high-frequency order book data for 30 stocks traded on Euronext Amsterdam covering a sample period over two months in 2008. We find strong evidence for the existence of common stochastic components in quotes and corresponding depths resulting in cointegration relations which significantly deviate from the bid-ask spread. In this sense, our results shed some light on the strength of co-movements in ask and bid prices depending on the underlying depth. Indeed, it turns out that order book inventory is highly persistent and reveals high-frequency dynamics resembling (near-)unit-root behavior. We show that incoming limit orders have significant impacts on subsequent ask and bid processes. It turns out that the magnitude and direction of quote adjustments strongly depend on the order's aggressiveness, its (relative) size and the prevailing depth in the book. In particular, we show the following results: (i) Quote adjustments are the stronger and the faster, the closer the incoming order is posted to the market. Most significant effects are reported for orders posted on up to two levels behind the market. For less aggressive orders, virtually no effects can be quantified. (ii) Limit orders temporarily narrow the spread. Converse effects are shown for market orders. In the long-run, these effects are reverted back in an asymmetric way. (iii) Large limit orders posted inside of the spread induce severe long-run effects pushing the market in the intended trading direction. In contrast, small limit orders posted inside of the spread tend to be picked up quickly inducing adverse price reactions. (iv) The long run market impact of aggressive market orders walking through the book is the higher the smaller the prevailing depth behind the market. (v) The effects are qualitatively stable across the market, where the absolute magnitudes of price impacts differ in dependence of underlying stock-specific characteristics. It turns out that approximately 60%-80% of the cross-sectional variation in market impacts can be explained by the trading frequency and the minimum tick size.

The remainder of this chapter is structured as follows. In Section 1.2, we describe the trading structure of Euronext Amsterdam and provide descriptive statistics. The econometric approach is explained in Section 1.3. Section 1.4 gives the estimation results and Section 1.5 provides the quantified price impacts of different types of limit orders. Finally, Section 1.6 concludes.

1.2 Data and Market Environment

Euronext is a purely electric limit order book market with price and time order precedence. During the continuous trading period between 9:00 and 17:30 CET, limit orders are submitted to a centralized computer system where they are matched to prevailing standing limit orders on the opposite side. If there is no match or the matched volume in the system is insufficient to exhaust the incoming order, the remaining order volume is placed in the order book. Euronext supports various order types like pure market orders (immediate order execution without a price limit), stop orders (automatic issuing of limit orders or market orders when a given price is reached), fill-or-kill (FOK) orders or iceberg orders.

Our dataset comprises limit order book (LOB) data of the 30 most frequently traded stocks at Euronext Amsterdam between August 1st and September 30th, 2008. Since on September 1st, Euronext changed the minimum tick size for some stocks, we analyze the two months August and September separately. This allows us to study the robustness of our findings under changing market conditions. Since these two months represent a generally turbulent market period, we further robustify our findings by replicating our analysis for a period which is less volatile. As we obtain quantitatively similar results, our findings can be seen as representative for different market conditions.[2]

The data contains information on the prevailing market depth (in terms of the number of shares) for the five best quotes on both sides of the market. Every trade and change of the order book are recorded in milliseconds. Preliminary analyzes (which are also supported by the findings given in Section 1.5) show that aggressive limit orders placed close to the best ask and bid have the highest market impact while induced price effects significantly decline with the distance to the spread. Accordingly, we focus only

[2]These results are not shown in this chapter but are available on our web appendix or upon request.

on the best three price levels in the book. Unlike the trade data which is well filtered by built-in filters in the database[3], the order book data is completely raw. We remove observations where (i) the spread is zero or negative, and (ii) ask or bid quotes change by more than 2%.[4] Moreover, to remove effects due to the opening and closing of the market, we discard data of the first five and last five minutes of the continuous trading period.

Matching of trade and LOB data is achieved by a matching algorithm which is described in details in Appendix A.1. This algorithm matches a trade with the corresponding LOB observation by comparing its price and volume with the resulting changes of quotes and depths in the book within an adaptively chosen time window. It minimizes the probability of misclassifications and as a by-product provides an estimate of the time asynchronicity between trade and LOB records.[5] To classify the initiation type of trades, we use a hybrid procedure according to Lee and Ready (1991). Firstly, we determine the type of trades which are located in more than one second time distance to previous trades using the mid-quote method. I.e., if a trade occurs with a price greater (less) than the most current mid-quote, it is classified as a buy (sell). If the trade price equals the mid-quote, it is marked as 'undetermined'. Secondly, 'undetermined' trades and trades which follow previous trades in less than one second time distance are classified by the tick-test method. Accordingly, if the trade price is higher (lower) than the previous one, it is identified as a buy (sell). If it does not change the price, it is categorized as the same type as the previous one. Finally, we identify sub-trades arising from the execution of a big market order against several (smaller) limit orders if they occur in less than one second after the previous trade and have the same initiation types. All corresponding sub-trades are consolidated to a single trade.

Table 1.1 gives descriptive statistics of the resulting August data used in this chapter.[6] We observe significantly more limit order activities than market orders. The average bid-ask spread is decreasing with the liquidity of the underlying stock. On average, second level market depth is higher than first level depth while it is approximately equal to the depth on the third level.

---

[3]Besides recording errors, block trades and trades in auction periods are excluded.

[4]In order to limit the volatility, Euronext NSC suspends continuous trading if prices change by more than 2%. This is not exactly the same rule as that implemented here, but it is reasonably mimicked.

[5]Due to technological progress in the last decades, the time delay between trade and quote records is nowadays hardly greater than one second. Consequently, the 'five-second' rule according to Lee and Ready (1991), which has been commonly used in empirical market microstructure literature, is not appropriate anymore for more recent datasets.

[6]Due to the aforementioned change in the minimum tick size, it is not appropriate to present joint summary statistics for both months. However, as the descriptive statistics for September are very similar to that for August, we do not present them here.

Table 1.1
Summary of synchronized trade and order book data.

The sample consists of the 30 most frequently traded stocks on Euronext Amsterdam. Market depth is measured in thousand shares. L1-L3 denote the order book level one to three. The period is from 1st to 31st August 2008.

stocks	#trades	#LO activ.	Ask			Bid			Mean of ask depth			Mean of bid depth		
	per day	per day	min	mean	max	min	mean	max	L1	L2	L3	L1	L2	L3
ING	1606.8	66569.1	20.255	21.518	23.290	20.250	21.507	23.275	3.64	3.94	4.12	3.45	3.90	4.14
FOR	1304.6	27574.0	8.770	9.351	10.160	8.760	9.338	10.150	16.78	25.76	25.03	16.35	26.25	24.20
RDSa	1166.2	48630.6	21.900	22.991	23.935	21.890	22.981	23.930	4.30	5.21	5.80	4.00	5.06	5.59
UNc	1152.1	46023.7	17.110	18.635	19.670	17.100	18.625	19.660	4.76	5.24	6.44	4.52	5.33	6.49
AHLN	1119.4	18730.3	7.540	8.510	8.970	7.530	8.502	8.960	7.89	9.80	10.23	8.18	10.64	10.59
PHG	1108.3	34722.0	20.875	22.381	23.465	20.870	22.368	23.450	2.18	2.36	2.70	1.95	2.19	2.59
AEGN	982.5	43270.2	7.290	7.909	8.400	7.280	7.902	8.395	5.12	4.99	4.86	4.98	4.98	4.79
AKZO	960.0	20061.2	35.460	39.571	41.920	35.400	39.541	41.910	0.89	0.96	1.00	0.78	0.90	0.98
KPN	954.0	20733.8	10.915	11.274	11.680	10.905	11.266	11.670	9.61	12.10	12.77	8.79	10.57	11.57
TNT	949.7	20412.7	22.040	24.598	27.000	22.030	24.566	26.970	1.57	1.91	2.15	1.51	1.96	2.24
HEIN	927.2	19782.1	29.540	31.796	33.660	29.520	31.767	33.600	0.98	1.10	1.13	0.92	1.00	1.04
ISPA	903.1	35708.2	49.990	52.694	56.440	49.910	52.661	56.420	1.85	2.76	3.66	1.97	3.08	3.84
ASML	853.8	26249.5	14.290	15.964	17.400	14.280	15.949	17.390	3.80	5.86	6.50	3.48	5.21	6.01
DSMN	826.7	21574.5	36.050	37.919	40.000	36.020	37.886	39.990	0.77	0.87	0.99	0.77	0.88	0.99
SBMO	603.7	18676.3	13.530	14.934	16.700	13.520	14.911	16.680	1.84	2.63	2.99	1.76	2.51	2.79
TOM2	505.3	16822.0	14.340	16.017	17.550	14.300	15.987	17.540	1.31	1.71	2.06	1.25	1.69	1.75
FUGRc	505.0	8846.5	43.620	47.701	53.200	43.610	47.631	53.180	0.56	0.54	0.52	0.49	0.49	0.47
WLSNc	548.8	16003.6	14.610	15.973	17.020	14.550	15.950	17.000	1.92	1.88	1.96	1.94	1.83	1.89
RAND	543.4	17265.2	17.710	19.432	21.430	17.690	19.397	21.400	1.09	1.56	1.75	1.07	1.47	1.47
ELSN	488.5	29702.2	10.390	11.049	11.510	10.350	11.035	11.500	7.27	11.57	11.96	6.81	11.34	12.44
BOSN	419.6	8013.0	32.320	36.323	41.900	32.250	36.247	41.890	0.52	0.52	0.49	0.53	0.51	0.47
BAMN	416.8	6334.1	9.900	10.736	12.220	9.860	10.714	12.200	2.06	2.35	2.38	1.99	2.25	2.19
SR	347.5	6396.6	10.370	11.588	13.200	10.360	11.563	13.180	1.70	1.80	1.76	1.72	1.71	1.48
CSMNc	340.2	7478.4	17.910	20.395	24.260	17.890	20.361	24.240	0.81	0.88	0.92	0.84	0.90	0.91
COR	327.1	12103.2	47.090	49.273	51.210	47.010	49.175	51.140	0.43	0.41	0.37	0.39	0.38	0.34
IMUN	292.7	5735.9	14.300	16.178	17.710	14.280	16.148	17.700	0.92	1.17	1.24	0.85	0.91	0.88
SMTNc	272.4	7648.8	43.920	52.282	60.440	43.840	52.112	60.300	0.22	0.25	0.22	0.26	0.27	0.26
NUTR	256.6	8043.2	41.160	43.275	44.900	41.120	43.192	44.890	0.40	0.36	0.33	0.37	0.38	0.38
USGP	248.5	6342.3	9.670	11.198	12.630	9.650	11.168	12.600	1.47	1.51	1.41	1.59	1.39	1.19
HEIO	181.0	14011.0	27.120	29.854	31.300	27.080	29.809	31.290	0.44	0.53	0.61	0.50	0.64	0.70

Table 1.2
Variable definitions
Events include limit order submissions, executions and cancellations. The market depth refers to the pending volume at the ordered available price levels in the LOB.

Variable	Description
p_t^a	logarithm of the best ask after the t-th event.
p_t^b	logarithm of the best bid after the t-th event.
$v_t^{a,l}$	logarithm of market depth at the l-th best ask after the t-th event.
$v_t^{b,l}$	logarithm of market depth at the l-th best bid after the t-th event.
BUY_t	dummy equal to one if the t-th event is a buyer-initiated trade.
$SELL_t$	dummy equal to one if the t-th event is a seller-initiated trade.

1.3 Econometric Modelling

1.3.1 A Cointegrated VAR Model for Quotes and Depths

Denote t as a (business) time index, indicating all order book activities, i.e., incoming limit or market orders as well as limit order cancellations. Then, p_t^a and p_t^b denote the best log ask and bid quotes instantaneously after the t-th order activity and $v_t^{a,j}$ and $v_t^{b,j}$ for $j = 1, \ldots, k$, define the log depth on the j-th best observed quote level on the ask and bid side, respectively. Furthermore, we introduce two dummy variables, BUY_t and $SELL_t$ indicating the occurrence of buy and sell trades, respectively. The inclusion of these two variables is necessary to distinguish between the effects caused by a market order and that induced by a cancellation. Both events remove volume from the book, however, presumably have quite different long run market impacts. Table 1.2 gives a detailed description of the variables.

To capture the high-frequency dynamics in quotes and depths we define a $K = (4 + 2 \times k)$-dimensional vector of endogenous variables

$$y_t := [p_t^a, p_t^b, v_t^{a,1}, \ldots, v_t^{a,k}, v^{b,1}, \ldots, v_t^{b,k}, BUY_t, SELL_t]'.$$

Note that the quote levels associated with $v_t^{a,j}$ and $v_t^{b,j}$ are not observed on a *fixed* grid at and behind the best quotes. Hence, their price distance to p_t^a and p_t^b is not necessarily exactly $j - 1$ ticks but might be higher if there are no limit orders on all possible intermediate price levels behind the market. To capture such 'gaps' in the order book, we could also include the limit prices associated with each order level posted behind the market and thus correspondingly extend the vector y_t. However, we decided to disregard this information because of two reasons. Firstly, Hautsch and Huang (2012a) show that trades "walking through the book", i.e., trades absorbing more than one price level in the limit order book occur extremely rarely for liquid stocks. Secondly, in liquid markets, the tick levels close to the best quotes are indeed mostly filled such that limit prices are on a fixed grid with constant distance to the corresponding best quotes. Consequently, we expect that the inclusion of all individual limit prices does not provide any additional information but just increases the dimension of the system. Finally, modelling log volumes instead of plain volumes is a common practice in many empirical studies to reduce the impact of extraordinarily large volumes. This is also suggested by Potters and Bouchaud (2003) studying the statistical properties of market impacts of trades. Moreover, using logs implies that changes in market depth can be interpreted as *relative* changes with respect to the current depth level.

Hence, we model log quotes, log depths and trading indicators as a restricted cointegrated vector autoregressive model of the order p (VAR(p)) with the vector error correction (VEC) form

$$\Delta y_t = \mu + \alpha\beta' y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + u_t, \tag{1.1}$$

where u_t is white noise with covariance matrix Σ_u, μ is a constant, Γ_i with $i = 1, \ldots, p-1$ is a $K \times K$ parameter matrix, α and β denote the $K \times r$ loading and cointegrating matrices with $r < K$. As we can safely assume that the trading indicators BUY_t and $SELL_t$ are stationary, we restrict the two first columns of β as $\beta_1 = [0, \ldots, 0, 1, 0]'$ and $\beta_2 = [0, \ldots, 0, 0, 1]'$.

For the impulse-response analysis below, it turns out to be more convenient to work with the reduced VAR representation in terms of the level of y_t,

$$y_t = \mu + \sum_{i=1}^{p} A_i y_{t-i} + u_t, \tag{1.2}$$

where $A_1 := I_K + \alpha\beta' + \Gamma_1$ with I_K denoting a $K \times K$ identity matrix, $A_i := \Gamma_i - \Gamma_{i-1}$ with $1 < i < p$ and $A_p := -\Gamma_{p-1}$.

We estimate the model (1.1) by the Full Information Maximum Likelihood (FIML) estimator proposed by Johansen (1991) and Johansen and Juselius (1990). Then, following Lütkepohl and Reimers (1992), we transform these estimates to representation (1.2). The corresponding procedure is shown in Appendix A.2 and A.3. By imposing the stationarity restrictions β_1 and β_2, all elements in the other cointegrating vectors associated with BUY_t and $SELL_t$ are automatically set to zero. This is guaranteed by the orthogonality among the estimated cointegrating vectors implied by FIML.

Note that market depth enters the vector y_t in levels and thus is treated as a possibly non-stationary variable. Though this is counter-intuitive for the behavior of depth over longer horizons, it is a reasonable assumption if depth is observed on very high frequencies. Moreover, modelling both quotes *and* depth in terms of a cointegration system guarantees consistency of parameter estimates irrespective of the possible (non-)stationarity of order book depth. Even if depth is truly stationary (and thus just corresponds to a (spurious) cointegration relation for itself), FIML estimates are consistent (though obviously not efficient).[7] Since we employ a high number of observations,

[7]See, for instance, Example 3.1 in Johansen (1995) for an illustration of this argument.

the possible loss of efficiency due to the neglect of a (stationarity) restriction is not very harmful in our context. If, however, we impose stationarity of depth and correspondingly restrict the cointegration vectors, we run the risk of producing inconsistent estimates if the restriction does not hold. Indeed, unit root tests applied in Section 1.4.1 indicate that the assumption of a unit root in depth observed on high frequencies cannot be rejected for many stocks. These arguments support the usefulness of a more robust statistical inference in form of an unrestricted cointegration system.

Model (1.2) can be further rotated in order to represent dynamics in spreads, relative spread changes, midquotes, midquote returns as well as (ask-bid) depth imbalances. Hence, the model is sufficiently flexible to capture the high-frequency dynamics of all relevant trading variables.[8]

Finally, in models involving only quote dynamics (e.g. Engle and Patton, 2004) or spread dynamics (e.g. Lo and Sapp, 2006), the error correction term $\beta' y_t$ is typically assumed to be equal to the spread implying a linear restriction $R'\beta = 0$ with $R' = [1, 1, 0, \ldots, 0]$. However, given the potential non-stationarity of order book depth, we do *not* impose this assumption here. As depth might contain information on the equilibrium (long run) state of the order book as well, we expect the existence of cointegration relations differing from spreads and involving both quotes *and* depths. As shown in the remainder of this chapter, this notion is actually supported by the data.

1.3.2 Limit Orders as Shocks to the System

In this section, we illustrate how to represent incoming orders as shocks to the system specified in equation (1.2). Whenever an order enters the order book, it (i) will change the depth in the book, (ii) may change the best quotes depending on which position in the queue it is placed, and (iii) will change the trading indicator dummy in case of a market order. We represent these changes in terms of an impulse vector $\delta := [\delta'_v, \delta'_p, \delta'_d]'$ with δ_v being a $2k \times 1$ vector associated with shocks to the depths, δ_p denoting a 2×1 vector consisting of shocks to the quotes and δ_d being a 2×1 vector representing shocks to the trading indicator dummy.

We design impulse response vectors associated with five scenarios commonly faced by market participants. As graphically illustrated by Figures 1.1 to 1.4, a three-level order book is initialized by the best ask $p_t^a = 1002$, best bid $p_t^b = 1000$, second best ask 1003, second best bid 999, and levels of depths on the bid side $V_t^{b,1} = 1$, $V_t^{b,2} = 1.5$, $V_t^{b,3} = V_t^{b,4} = 1.4$. The following scenarios are considered:[9]

Scenario 1a (normal limit order): Arrival of a buy limit order with price 1000 and size 0.5 to be placed *at the market*. As shown in Figure 1.1, this order will be

[8]Note that we do not impose an explicit constraint ensuring the positiveness of bid-ask spreads. As shown on the companion website, this restriction is implicitly satisfied by our estimates in virtually all cases.

[9]For sake of brevity, the scenarios are only characterized for buy orders. For sell orders, the setting is correspondingly adapted to the other side of the market.

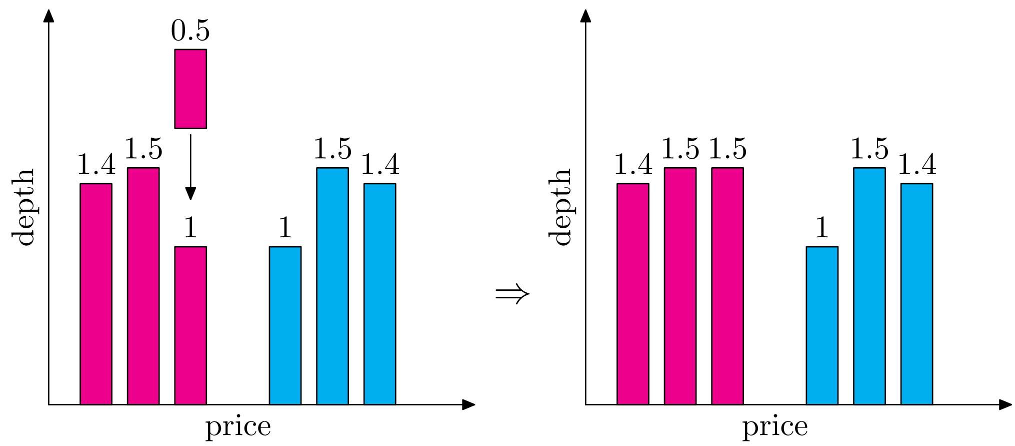


Figure 1.1 (Scenario 1a (normal limit order)): An incoming buy limit order with price 1000 and size 0.5. It affects only the depth at the best bid without changing the prevailing quotes or resulting in a trade.

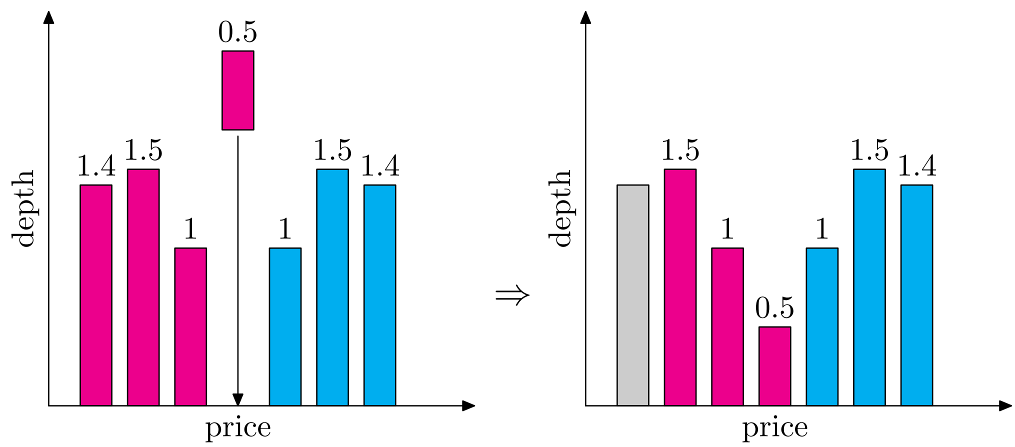


Figure 1.2 (Scenario 2 (aggressive limit order)): An incoming buy limit order with price 1001 and size 0.5 improving the best bid and changing all depth levels on the bid side of the order book.

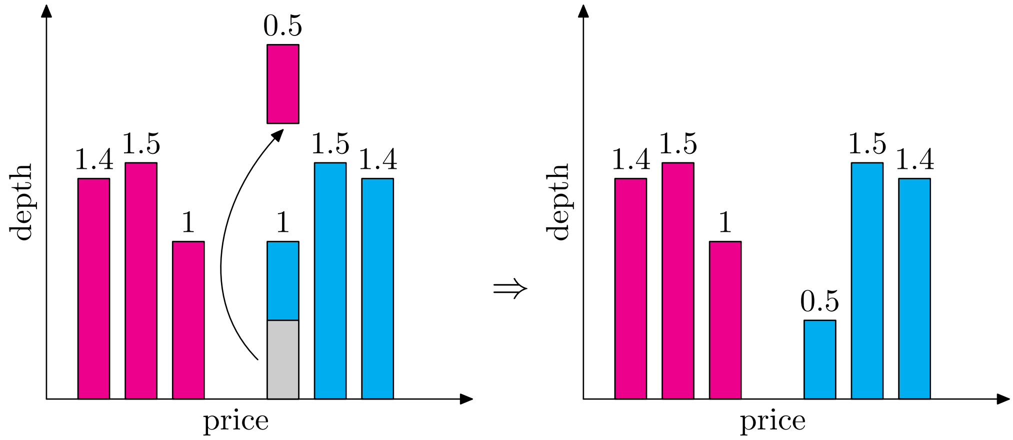


Figure 1.3 (Scenario 3 (normal market order)): An incoming buy market order with price 1002 and size 0.5 which results in a buyer-initiated (BUY) trade.

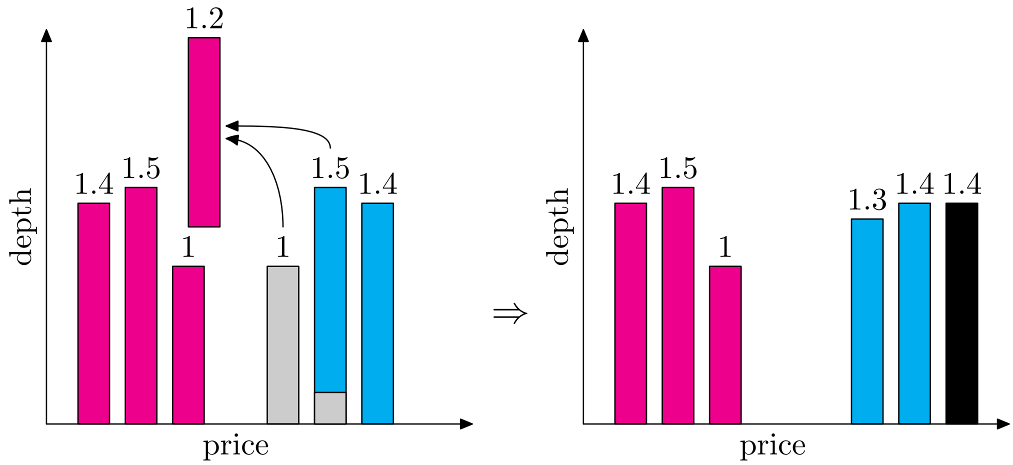


Figure 1.4 (Scenario 4 (aggressive market order)): An incoming buy market order with price 1003 and size 1.2 'walking through' the order book and simultaneously changing all depth levels on the ask side.

consolidated at the best bid without changing the prevailing quotes. Because the initial depth on the first level is 1.0, the change of the log depth is $\ln(1.5) \approx 0.4$. Correspondingly, the shock vectors are given by $\delta_v = [0, 0, 0, 0.4, 0, 0]'$, $\delta_p = \delta_d = [0, 0]'$.

Scenario 1b (passive limit order): Arrival of a buy limit order with price 999 and size 0.5 to be posted *behind the market*. As in the scenario above, it does not change the prevailing quotes and only affects the depth at the second best bid. We have $\delta_v = [0, 0, 0, 0, \ln(2) - \ln(1.5) \approx 0.29, 0]'$, $\delta_p = \delta_d = [0, 0]'$.

Scenario 2 (aggressive limit order): Arrival of a buy limit order with price 1001 and size 0.5 to be posted inside of the current spread. Figure 1.2 shows that it improves the best bid by 0.1% and accordingly shifts all depth levels on the bid side. The resulting shock vector is given by $\delta_v = [0, 0, 0, (\ln(0.5) \approx -0.69), (\ln(1/1.5) \approx -0.4), (\ln(1.5/1.4) \approx 0.07)]'$, $\delta_p = [0, 0.001]'$ and $\delta_d = [0, 0]'$.

Scenario 3 (normal market order): Arrival of a buy order with price 1002 and size 0.5. This order will be executed immediately against standing limit orders at the best ask. Because it absorbs liquidity from the book, it shocks the corresponding depth levels negatively. Figure 1.3 depicts the corresponding changes of the order book as represented by $\delta_v = [\ln(0.5) \approx -0.69, 0, 0, 0, 0, 0]'$, $\delta_p = [0, 0]'$ and $\delta_d = [1, 0]'$.

Scenario 4 (aggressive market order): Arrival of a buy order with price 1003 and size 1.2. It 'walks up' the order book. As shown in Figure 1.4, the best ask quote and all depth levels are simultaneously shifted resulting in the shock vector $\delta_v = [(\ln(1.3) \approx 0.26), (\ln(1.4/1.5) \approx -0.07), 0, 0, 0, 0]'$, $\delta_p = [(1/1002) \approx 0.001, 0]'$ and $\delta_d = [1, 0]'$.

Table 1.3 summarizes the shock vectors implied by the illustrated scenarios.

1.3.3 Measuring the Market Impact

We quantify the market impact of limit orders as the implied expected short-run and long-run shifts of the ask and bid after their submissions. This reaction is captured by the impulse response function,

$$f(h; \delta_y) = \mathrm{E}[y_{t+h}|y_t + \delta_y, y_{t-1}, \cdots] - \mathrm{E}[y_{t+h}|y_t, y_{t-1}, \cdots], \tag{1.3}$$

where the shock on quotes, depths and trading indicators is denoted by $\delta_y := [\delta_p', \delta_v', \delta_d']'$ and h is the number of periods (measured in 'order event time').

Note that we do not have to orthogonalize the impulse since contemporaneous relationships between quotes and depths are captured by construction of the shock vector. Moreover, our data is based on the arrival time of orders avoiding time aggregation as another source of mutual dependence in high-frequency order book data.

Table 1.3
Shock vectors implied by the underlying five scenarios

Initial order book: best ask $p_t^a = 1002$, best bid $p_t^b = 1000$, second best ask = 1003, second best bid = 999. Volumes on the ask/bid side: $V_t^{a/b,1} = 1$ at the best bid, $V_t^{a/b,2} = 1.5$ at the second best bid, and $V_t^{a/b,3} = V_t^{a/b,4} = 1.4$ at the third and fourth best bids, respectively. Notation: δ_v denotes shocks on market depths; δ_p denotes shocks on the best bid and best ask; δ_d denotes shocks on trading indicator variables.

Scenario	limit order	shock vectors		
	(dir,price,size)	δ'_v	δ'_p	δ'_d
'normal limit order'	(Bid,1000, 0.5)	$[0, 0, 0, 0.4, 0, 0]$	$[0, 0]$	$[0, 0]$
'passive limit order'	(Bid,999, 0.5)	$[0, 0, 0, 0, 0.29, 0]$	$[0, 0]$	$[0, 0]$
'aggressive limit order'	(Bid,1001, 0.5)	$[0, 0, 0, -0.69, -0.4, 0.07]$	$[0, 0.001]$	$[0, 0]$
'normal market order'	(Bid,1002, 0.5)	$[-0.69, 0, 0, 0, 0, 0]$	$[0, 0]$	$[1, 0]$
'aggressive market order'	(Bid,1003, 1.2)	$[0.26, -0.07, 0, 0, 0, 0]$	$[0.001, 0]$	$[1, 0]$

Using impulse-response analysis to retrieve the market impact has two major advantages. First, in contrast to an analysis of estimated VEC coefficients which only reveals the immediate impact, it enables us to examine both long-run and short-run effects. Second, it allows us to straightforwardly quantify the joint effect induced by simultaneous changes of several variables given a certain state of other variables.

We consider two moving average (MA) representations of the cointegrated VAR model. The first one is based on the reduced form given by equation (1.2). This representation allows us to compute the path of the response function over time. The second one is the Granger representation based on the VECM form in equation (1.1) which enables us to explicitly compute the permanent (long-run) response.

We start our discussion with the first MA representation. The companion VAR(1) form of the VAR(p) model in equation (1.2) is given by

$$Y_t = \boldsymbol{\mu} + \mathbf{A}Y_{t-1} + U_t, \tag{1.4}$$

where

$$\boldsymbol{\mu} := \underbrace{\begin{bmatrix} \mu \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{Kp\times 1}, \; Y_t := \underbrace{\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}}_{Kp\times 1}, \; U_t := \underbrace{\begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{Kp\times 1}$$

and

$$\mathbf{A} := \underbrace{\begin{bmatrix} A_1 & \cdots & A_{p-1} & A_p \\ I_K & & 0 & 0 \\ & \ddots & \vdots & \vdots \\ 0 & \cdots & I_K & 0 \end{bmatrix}}_{Kp \times Kp}.$$

Successively substituting Y yields

$$Y_t = M_t + \sum_{i=0}^{t-1} \mathbf{A}^i U_{t-i}, \tag{1.5}$$

where $M_t = \mathbf{A}^t Y_0 + \sum_{i=0}^{t} \mathbf{A}^i \boldsymbol{\mu}$ consists of terms of an initial value and a deterministic trend, which are irrelevant for the impulse-response analysis. Let $J := [I_K : 0 : \cdots : 0]$ be a $K \times Kp$ selection matrix with $JY_t = y_t$. Pre-multiplying J on both sides of equation (1.5) and using $U_t = J' u_t$ gives

$$y_t = JM_t + \sum_{i=0}^{t-1} J\mathbf{A}^i J' u_{t-i}. \tag{1.6}$$

Then, the linear impulse-response function according to equation (1.3) can be written as

$$f(h; \delta_y) = J\mathbf{A}^h J' \delta_y. \tag{1.7}$$

Given the consistent estimator $\widehat{a}$ for $a := \mathrm{vec}(A_1, \ldots, A_p)$ in equation (1.2),

$$\sqrt{T}(\widehat{a} - a) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widehat{a}}),$$

Lütkepohl (1990) shows that the asymptotic distribution of the impulse-response function is given by

$$\sqrt{T}(\hat{f} - f) \xrightarrow{d} \mathcal{N}(0, G_h \Sigma_{\widehat{a}} G_h'), \tag{1.8}$$

where $G_h := \partial\, \mathrm{vec}(f) / \partial\, \mathrm{vec}(A_1, \cdots, A_p)'$. This expression can be explicitly written as

$$G_h = \sum_{i=0}^{h-1} \left(\delta_y' J (\mathbf{A}')^{h-1-i} \otimes J\mathbf{A}^i J' \right). \tag{1.9}$$

In order to compute the long-run effect, we apply Granger's Representation Theorem to model (1.1) yielding

$$y_t = C \sum_{i=1}^{t} (u_i + \mu) + C_1(L)\,(u_t + \mu) + V, \tag{1.10}$$

where

$$C = \beta_\perp \left(\alpha_\perp' \left(I_K - \sum_{i=1}^{p-1} \Gamma_i \right) \beta_\perp \right)^{-1} \alpha_\perp' \,. \tag{1.11}$$

Here, L is the lag operator and the power series $C_1(z)$ is convergent for $|z| < 1 + \xi$ for some $\xi > 0$. V depends on initial values, such that $\beta' V = 0$. The Granger representation decomposes the cointegrated process into a random walk term (C term), a stationary process (C_1 term) and a deterministic term (V). Because of the convergence of the series $C_1(z)$, the response implied by this sub-process will be zero in the long run. Moreover, the deterministic term V is irrelevant for the impulse response. Therefore, the permanent response of the system is completely determined by the first term. Note that the shock δ_y causes this term changing by $C\delta_y$. Thus, we can express the permanent response as

$$\bar{f}(\delta_y) := \lim_{h \to \infty} f(h; \delta_y) = C\delta_y. \tag{1.12}$$

Note that given α and β, $\alpha_\perp$ and $\beta_\perp$ are not uniquely identified. However, the right hand side of equation (1.11) is invariant with respect to the choice of these bases. Therefore, $\bar{f}(\delta_y)$ is unique given the parameters and the shock vector in model (1.1). In practice, estimated responses and their covariances are obtained by replacing the unknown parameters in equation (1.7), (1.8) and (1.12) by their estimates.

1.4 Estimation Results

The underlying order book data contains bid and ask quotes as well as five levels of depth. Preliminary analyzes show that the depths on the fourth and fifth levels do not have significant effects on bid and ask quotes. Therefore, in our empirical study, we only use market depths up to the third level. In order to keep the analysis tractable, we reduce the computational burden induced by the high number of observations by separately estimating the model for each of the 43 trading days. This strategy allows us also to address possible structural changes, e.g., due to stock specific news announcements or overnight effects. The market impact is then computed as the monthly average of individual (daily) impulse responses. Likewise, confidence intervals are computed based on daily averages. To account for a structural break due to the change of the tick size for some stocks on September 1, 2008, we treat the two months August and September separately.

For sake of brevity we refrain from presenting all individual results for the 30 analyzed stocks in this chapter. We rather illustrate the analyzed effects for the stock Fortis (FOR in Table 1.1) in August 2008. Fortis is one of the most actively traded stocks and is representative for a major part of the market. The results for the remaining stocks and the remaining periods are provided in a web appendix on `http://amor.cms.hu-berlin.de/~huangrui/project/impact_of_orders`. As shown in the web appendix and discussed in more detail in Section 1.5.5, the effects are qualitatively remarkably similar across the market though the magnitudes of market impacts differ in dependence of underlying stock-specific characteristics.

The empirical analysis employs a VAR(15) specification which is selected based on residual diagnostics and information criteria. Testing for serial correlation using the Ljung-Box test according to Ljung and Box (1978) reveals almost no remaining serial

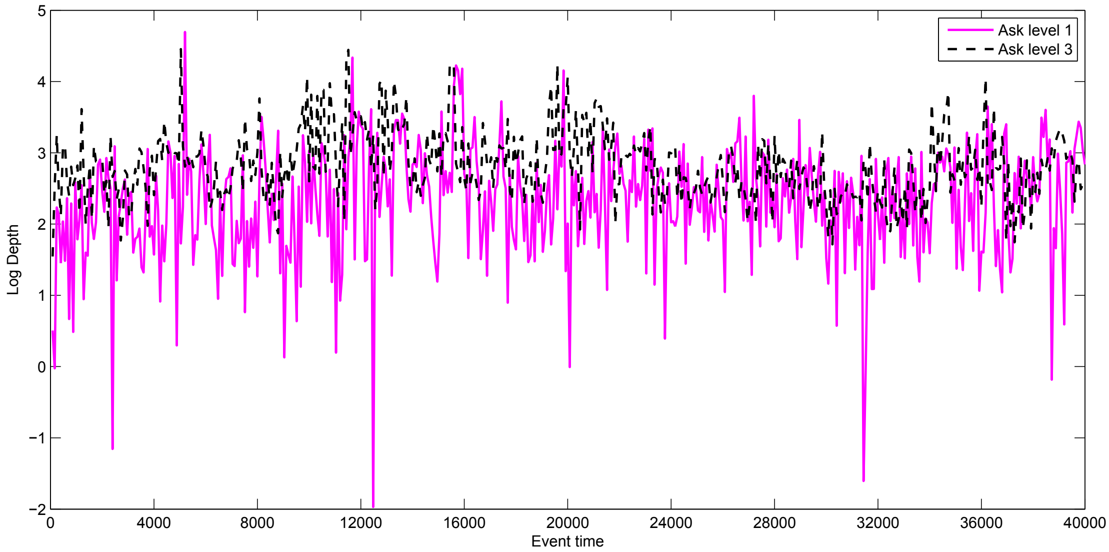


Figure 1.5: Time series plot of log market depths (measured in thousand share units). Trading of Fortis, Euronext, Amsterdam, August 1st, 2008.

correlation in the residuals for all regressions based on a 1% level using ten lags. The corresponding statistics are also recorded in the web appendix.

1.4.1 Statistical Properties of Market Depth

Figure 1.5 provides time series plots of depths on the best ask and third best ask level of the order book for a single (though representative) trading day for Fortis. A general finding is that the depth behind the market is typically greater than that at the market. Furthermore, there is evidence for co-movements between the individual depth levels, partially because of the 'shift' effect induced by aggressive orders, e.g., limit orders posted inside of spreads or market orders completely absorbing the best price levels.

Figure 1.6 depicts the unconditional distributions and autocorrelation functions of log market depth. We observe that the distribution of depth behind the market is similar, though they are quite different from those at the market. The same pattern is also observed for the autocorrelation functions. These empirical peculiarities are due to the fact that there is obviously more order activity at the market than behind the market. Consequently, market depth is more frequently changed at the best level inducing a lower persistence than at higher levels. This might also explain why the unconditional distribution of depth is more dispersed than that of depth behind the market.

Table 1.4 shows the results of Augmented Dickey-Fuller (ADF) and KPSS tests for quotes and market depth. While the quote series are obviously integrated, we obtain conflictive findings for the depth series. The ADF tests reject the null hypothesis of a unit root in first level depth in 83% of all cases (across stocks and days), whereas the

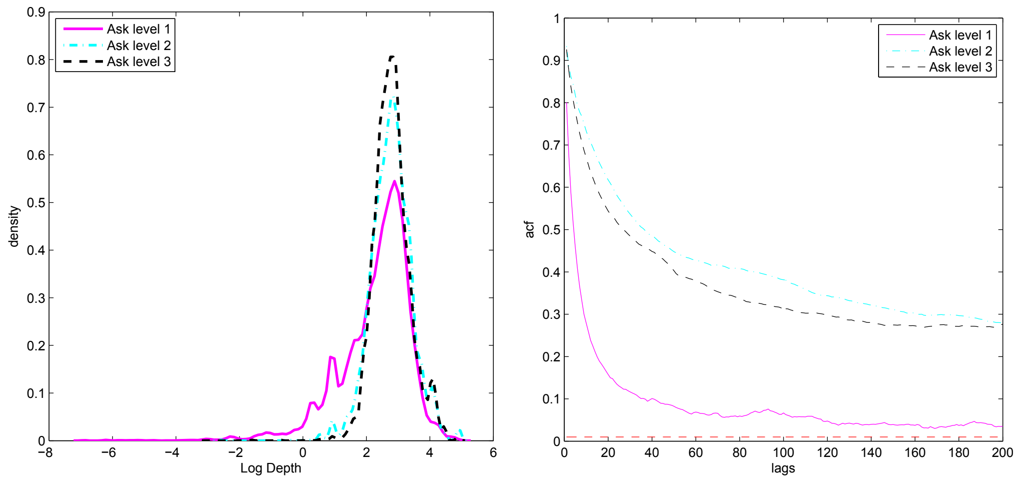


Figure 1.6: Left: Kernel density estimates of (log) market depths. **Right:** Autocorrelation functions of (log) market depths. Trading of Fortis, Euronext, Amsterdam, August 1st, 2008.

Table 1.4
Stationarity tests on quotes and market depths

Augmented Dickey-Fuller (ADF) tests and KPSS tests for the 30 selected stocks on each of the 43 trading days, i.e., 1290 time series for each variable. The chosen lag length is 50. The reported numbers are the sum of rejections at the 1%-level. In the ADF test, the null hypothesis is that there is an unit root in the process. In the KPSS test, the null hypothesis is that there is *no* unit root in the process.

Variables	p^a	p^b	$v^{a,1}$	$v^{a,2}$	$v^{a,3}$	$v^{b,1}$	$v^{b,2}$	$v^{b,3}$
ADF	8	4	1072	975	933	1087	975	949
(%)	(0.62)	(0.31)	(83.1)	(75.58)	(72.32)	(84.26)	(75.58)	(73.56)
KPSS	1284	1283	871	905	982	846	896	979
(%)	(99.53)	(99.45)	(67.51)	(70.15)	(76.12)	(65.58)	(69.45)	(75.89)

Table 1.5
Representative estimates of cointegrating vectors
The vectors are sorted according to their corresponding eigenvalues in Johansen's ML approach. The first two vectors are fixed to $\beta_1 = [0, \ldots, 0, 1, 0]$ and $\beta_2 = [0, \ldots, 0, 0, 1]$ representing stationary process of trading indicators. Correspondingly, all entries in $\widehat{\beta}_3$ to $\widehat{\beta}_9$ associated with the trading indicator variables, BUY and $SELL$, are set to zero and are omitted. Trading of Fortis at Euronext, Amesterdam on August 1, 2008.

Variable	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$	$\hat{\beta}_9$
p^a	-0.9987	-1.0000	1.0000	-0.9989	1.0000	-1.0000	0.9399
p^b	1.0000	0.9853	-0.9968	1.0000	-0.9767	0.7048	-1.0000
$v^{a,1}$	-0.0173	0.1629	-0.0416	0.0222	-0.0803	0.0919	-0.1039
$v^{a,2}$	0.0070	-0.0486	0.0322	-0.0839	-0.1915	0.5869	-0.6605
$v^{a,3}$	-0.0070	0.0140	-0.0212	0.0108	0.2980	0.6104	-0.5413
$v^{b,1}$	-0.0081	-0.1412	-0.0398	0.0827	-0.0442	-0.0807	-0.0933
$v^{b,2}$	0.0003	0.0527	0.0430	0.2321	0.0167	-0.8162	-0.4652
$v^{b,3}$	-0.0002	-0.0342	-0.0212	-0.2988	0.0796	-0.9414	-0.3337

KPSS tests reject the stationarity in 67% of all cases. For higher level depth, the evidence against stationarity in depth is even higher. As discussed in Section 1.3.1, we explain these findings by the fact that order book depth is an inventory variable which over short horizons is strongly autocorrelated and tend to behave like an I(1) process. On the other hand, aggressive trading and limit order arrival create fluctuations in depth which are less predictable and reduce the strong persistence over longer intervals. Extreme changes arise, for instance, whenever first level depth is absorbed by an incoming order or, alternatively, is undercut by an incoming aggressive limit order, and thus the entire order book is shifted. Hence, from this discussion and the empirical findings we can conclude that depth might naturally contain stationary and non-stationary components where the latter tend to dominate over very short horizons. Given these results, it is in any case recommended to model depth as a non-stationary variable within a cointegrated VAR framework. As discussed in Section 1.3.1, this proceeding ensures consistency of parameter estimates even if depth might be stationary and, e.g., is fractionally cointegrated (see Johansen and Nielsen, 2010).

1.4.2 Estimated Cointegration Relations

For sake of brevity, we refrain from showing the individual estimates of **A** and **B**. Nevertheless, it is interesting to highlight the estimated cointegration relations. According to Johansen's trace statistics we identify seven cointegration relations among quotes and depths. Table 1.5 shows the estimated cointegrating vectors for a representative trading

Table 1.6

Representative estimates of the loading matrix

Values in parentheses are $t-$statistics. Trading of Fortis at Euronext, Amesterdam on August 1, 2008.

	$\hat{\beta}_3^{'} y_{t-1}$	$\hat{\beta}_4^{'} y_{t-1}$	$\hat{\beta}_5^{'} y_{t-1}$	$\hat{\beta}_6^{'} y_{t-1}$	$\hat{\beta}_7^{'} y_{t-1}$	$\hat{\beta}_8^{'} y_{t-1}$	$\hat{\beta}_7^{'} y_{t-1}$
p_t^a	0.0818	-0.0104	-0.0084	0.0042	0.0022	-0.0004	0.0002
	(18.45)	(-2.34)	(-50.11)	(18.42)	(8.50)	(-2.65)	(0.89)
p_t^b	-0.0691	-0.0133	0.0026	0.0041	0.0007	-0.0004	-0.0000
	(-15.91)	(-3.07)	(15.58)	(18.33)	(2.77)	(-2.59)	(-0.07)
$v_t^{a,1}$	2.1666	-0.4160	0.5319	0.0108	0.1228	0.0013	0.0052
	(18.25)	(-3.50)	(118.70)	(1.76)	(17.52)	(0.31)	(0.75)
$v_t^{a,2}$	-0.2935	0.0512	-0.1776	0.0423	0.0995	-0.0074	0.0114
	(-7.44)	(1.29)	(-119.21)	(20.65)	(42.68)	(-5.52)	(5.01)
$v_t^{a,3}$	0.0589	-0.0117	0.1293	-0.0176	-0.1314	-0.0072	0.0079
	(1.58)	(-0.31)	(92.30)	(-9.14)	(-59.94)	(-5.72)	(3.71)
$v_t^{b,1}$	1.5157	0.4704	0.7016	-0.1435	0.0596	0.0017	0.0025
	(12.75)	(3.96)	(156.39)	(-23.26)	(8.49)	(0.42)	(0.37)
$v_t^{b,2}$	-0.2284	-0.0607	-0.2373	-0.1053	-0.0036	0.0087	0.0074
	(-5.99)	(-1.59)	(-165.07)	(-53.26)	(-1.58)	(6.72)	(3.40)
$v_t^{b,3}$	-0.0492	0.0060	0.1322	0.1253	-0.0176	0.0119	0.0034
	(-1.39)	(0.17)	(99.52)	(68.59)	(-8.44)	(10.01)	(1.66)

day, where we omit the two known cointegrating vectors associated with the (stationary) trading indicators. Likewise we also omit the corresponding entries in the remaining cointegrating vectors as they are zero by construction. The resulting vectors are ordered according to their corresponding eigenvalues reflecting their likelihood contributions. Table 1.6 shows the estimated loading matrix, $\hat{\alpha}$, and corresponding $t-$statistics. We observe that not quotes but also depth variables have a significant loading on most of the six cointegration relations.

Figure 1.7 depicts the time series of the estimated cointegration relations. The series are quite different from that of the bid-ask spread (i.e., the difference between ask and bid quotes) which would be expected if depth does not belong to the cointegration vector and is also depicted in the figure. Compared to the spread which reflects a very discrete behavior, the cointegration relations are much more smooth. Nevertheless, as *any* linear combination of these vectors results into a further cointegration relation, it is required to formally test whether the estimated cointegration relations are indeed different from the bid-ask spread. The corresponding likelihood ratio test of the null hypothesis $R'\beta = 0$ with $R = [1, 1, 0, \ldots, 0]'$ rejects at a 1% significance level for all regressions (except one) for Fortis. Hence, we obtain significant evidence for depth being part of the cointegration relations influencing long-term equilibria of quotes and depth.[10]

Interpreting the estimated cointegrating vectors, we can derive several implications. The first five cointegration relations are mostly linear combinations of spreads and depths. Specifically, the first one is quite similar to the bid-ask spread as the coefficients for the depth variables are comparably small. The second cointegration relation seems to involve the balance of at-the-market depth since the coefficients of $v^{a,1}$ and $v^{b,1}$ are similar in magnitude and opposite in sign. The most interesting relationships are implied by the last two cointegrating vectors revealing relatively large (and different) coefficients associated with depth. This indicates that depth has a significant impact on the long-term relationship between quotes. Intuitively, the connection between ask and bid quotes becomes weaker (and thus deviates from the spread) if the depth is less balanced between both sides of the market. Hence, depth has a significant impact on quote dynamics and should be explicitly taken into account in a model for quotes. These findings support the idea of a cointegration model for *both* quotes and depth.

1.5 Estimated Market Impact

1.5.1 Limit Orders Placed At or Behind the Market

Consider the impact of an incoming at-the-market limit order as described in Scenario 1 in Section 1.3.2. Figure 1.8 shows the impulse responses induced by buy and sell limit

[10]It is well known that likelihood ratio tests on cointegration vectors tend to be biased towards rejecting the null hypothesis too often in finite samples, see, e.g., Gredenhoff and Jacobson (2001) and Haug (2002). However, given the high number of observations used in our study, these effects should not be too strong.

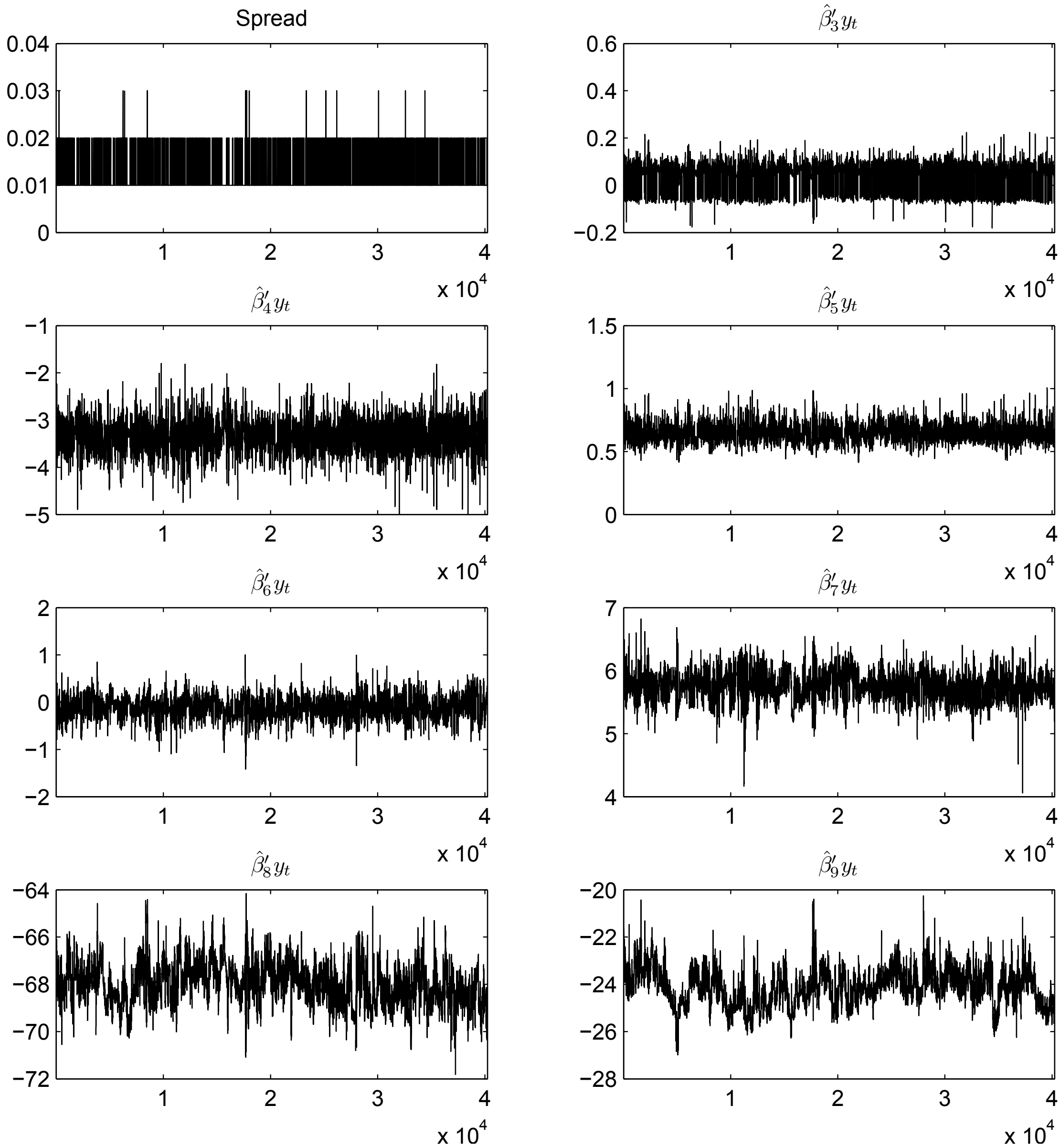


Figure 1.7: Time series of estimated cointegration relations. The corresponding cointegrating vectors are documented in Table 1.5. We suppress the two cointegrating relationships associated with the trading indicator series. Trading of Fortis at Euronext, Amsterdam, on August 1st, 2008.

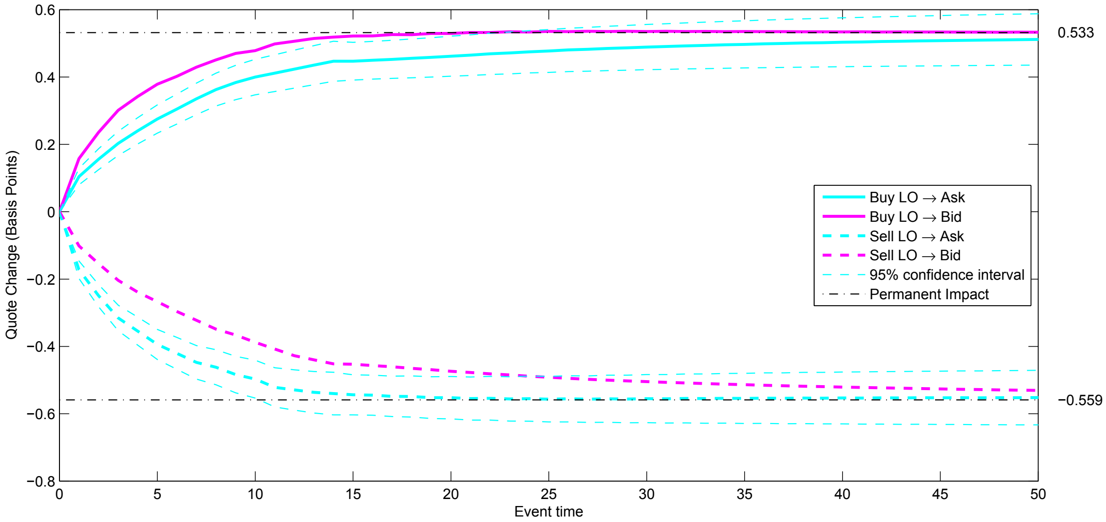


Figure 1.8: Changes of ask and bid quotes induced by buy/sell limit orders placed at the market (level one) with a size equal to the half of the depth on the first level. The marked number on the vertical axes indicates the magnitude of the permanent impact. The blue dotted lines indicate the corresponding 95%-confidence intervals. Trading of Fortis at Euronext, Amsterdam in August 2008. LO: limit order.

orders with a size equal to half of the depth at the best quotes.[11] The impulse response function starts at zero since such a limit order does not directly change the ask and bid. As expected, both ask and bid tend to significantly increase (decrease) after the arrival of a buy (sell) limit order. Induced by the cointegration setting, quotes converge to a (new) permanent level at which the information content of the incoming limit order is completely incorporated. The confidence intervals reflect that the shift is statistically highly significant.

We observe that quotes adjust relatively quickly reaching the new level after approximately 20 lags. Recall that time is measured in terms of limit order book activities. Hence, the adjustment speed measured in physical time ultimately depends on the underlying frequency of order activities and differs across the market. However, the fact that the speed of stock-specific quote adjustments (in terms of a 'limit order clock') is widely stable across the market, indicates that such a business time scale is appropriate for market-wide comparisons across stocks.

An interesting fact is that after the arrival of a buy limit order, the bid tends to increase more quickly than the ask. A reverse effect is observed after the arrival of a sell limit order. This asymmetry introduces a one-sided and temporary decrease of the bid-ask spread. We explain this phenomenon by the fact that traders observing

[11] In all figures illustrating impulse responses, the legend 'A → B' is interpreted to reflect 'the impact on B induced by A'.

an incoming limit order on the same side of the market tend to compete for provided liquidity by undercutting quotes. Moreover, the higher depth at the bid generates a (delayed) liquidity demand on the ask side shifting upward the ask as well. We thus refer this phenomenon to be a liquidity-motivated effect.

Our findings can be interpreted in terms of pure market mechanisms. The market equilibrium is perturbed by a limit order in two ways. On one hand, the limit order indicates an investor's willingness to buy or sell and thus increases the supply or demand of the underlying asset. The market price changes to incorporate this temporary imbalance of supply and demand. One the other hand, an incoming limit order increases the supply of liquidity in the market. A narrowing of the spread reduces transaction costs and causes a re-balancing of supply and demand of liquidity. See, e.g., the simulation study by Yamamoto (2011) on the effects of the state of the limit order book on investors' strategies.

The significant permanent impact induced by an incoming limit order indicates that it contributes to price discovery. Thus, market participants perceive that limit orders carry private information which is in contrast to the common assumption in theoretical literature that informed traders only take liquidity but do not provide it. On the other hand, it is supported by the experiment by Bloomfield, O'Hara, and Saar (2005) showing that informed traders use order strategies involving both market orders and limit orders to optimally capitalize their informational advantage and in line with Mike and Farmer (2008) and Chiarella, Iori, and Perelló (2009) suggesting that there is a link between the properties of order flow and those of prices.

Given the setting of the book we observe that a limit order increasing first level depth by 50% shifts quotes by 0.5 to 0.6 basis points. Though this effect is generally rather small, it is economically significant if the tick size is small. Moreover, note that the magnitude of the market impact is log-linear in the order size. In practice, a big limit order posted on a thin order book might affect the market much more strongly than in our scenarios.

In order to explore the role of the order's position in the book, Figure 1.9 depicts the bid prices' reactions induced by buy limit orders placed at the market (level one) and behind the market (level two and three). We observe a negative correlation between the magnitude of quote reactions and the orders' distance from the spread. The at-the-market limit order induces significantly faster market reactions than the behind-the-market limit order. Nonetheless, the long-term impact of level one and level two limit orders is only approximately 20% smaller. Hence, it turns out that behind-the-market orders can significantly shift the market though the quote adjustment is slower.[12] This result holds for level two orders and (to a weaker extent) for level three orders. However, for orders posted deeper in the book virtually no market impacts can be identified.

Eom, Lee, and Park (2009) find evidence that traders could have made extra profits using microstructure-based manipulations on the Korean Exchange (KRX) during a period between 2001 and 2002. In this period, KRX disclosed the total quantity on each

[12]In order to improve the graphical illustrations, we refrain from showing the corresponding confidence intervals. They are quite similar to those shown in Figure 1.8.

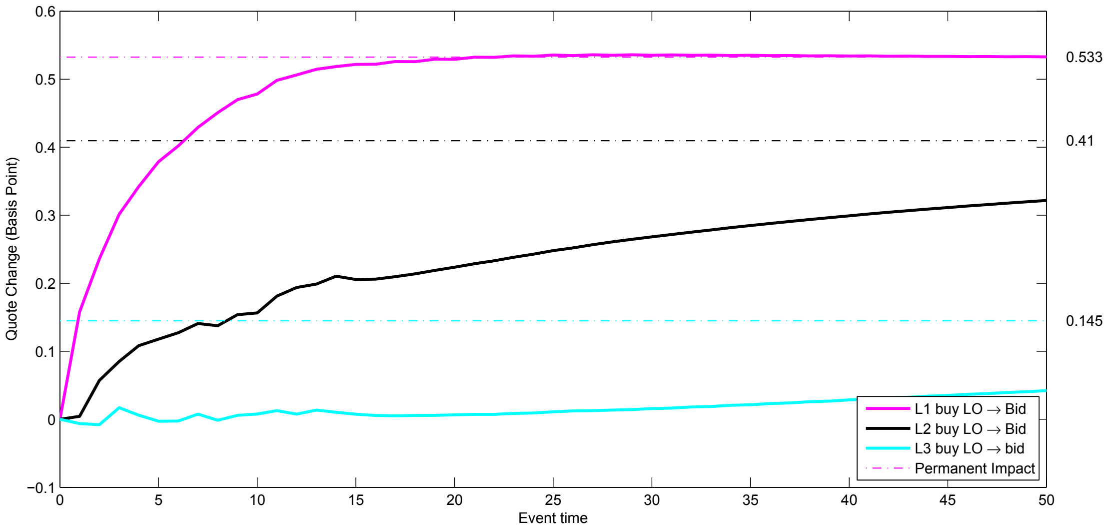


Figure 1.9: Changes of bid quotes induced by buy limit orders placed at the market (level one) and behind the market (level two and three). The order size equals to half of that at the best bid. The initial order book equals the corresponding monthly average shown in Table 1.1. The marked number on the vertical axes indicates the magnitude of the permanent impact. Trading of Fortis at Euronext, Amsterdam in August, 2008. L1: level one. L2: level two. L3: level three. LO: limit order.

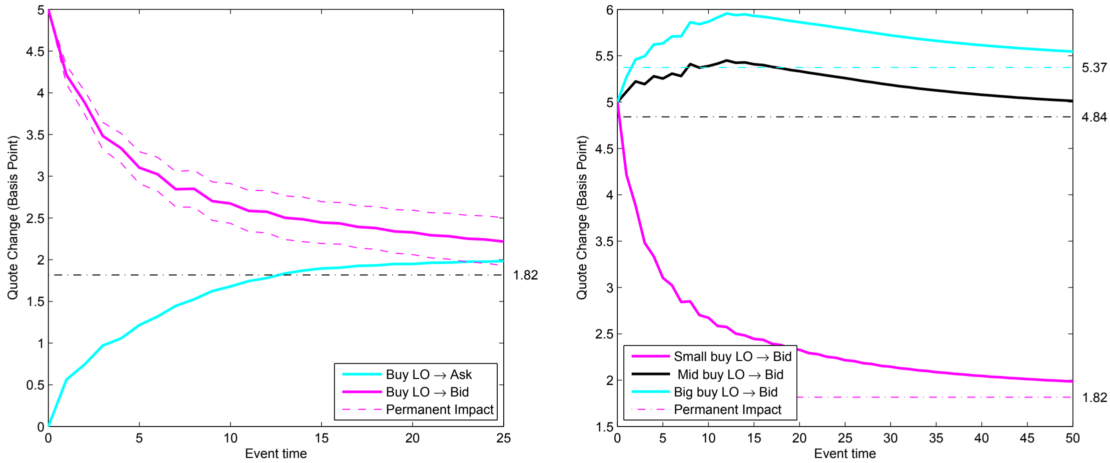


Figure 1.10: Left: Changes of quotes induced by buy limit orders placed inside of the spread with a size equal to the depth at the bid. **Right:** Changes of bid induced by buy limit orders placed inside of the spread with different sizes. The initial order book equals to the corresponding monthly average shown in Table 1.1. Small size: depth at the bid. Mid size: 7 times of the depth at the bid. Big size: 15 times of the depth at the bid. Trading of Fortis at Euronext, Amsterdam in August, 2008. LO: limit order.

side of LOB without fully disclosing the prices at which these orders have been placed. The manipulation strategy resulted in placing huge numbers of behind-the-market limit orders on the opposite side of the market inducing price moves in the favorite direction without having these orders executed. Our finding shows that this kind of manipulation is indeed possible. However, whether it is economically profitable in Euronext ultimately depends on (relative) order sizes. In order to move quotes in her favorite direction, the trader has to submit rather big limit orders close to the market. Then, she faces the risk that these orders may be picked up.

1.5.2 Limit Orders Placed Inside Of the Spread

Limit orders placed inside of the bid-ask spread perturb the LOB dynamics in a more complex way. Apart from providing liquidity to the order book, they directly improve the ask or bid. This quote adjustment induces a reduction of the spread, establishes a new best quote level and correspondingly shifts all depth levels on the corresponding side of the book upward (or downward, respectively). The system seeks the new equilibrium on a path recovering from the immediate quote change and simultaneously re-balancing liquidity. Given our setting, we assume that a buy limit order inside of the spread induces a 5 basis points increase of the bid. However, as shown in the left plot of Figure 1.10, the long-run price impact is just 1.8 basis points. The immediate quote movement is reverted back by approximately 65%. This is induced either by sell trades picking up

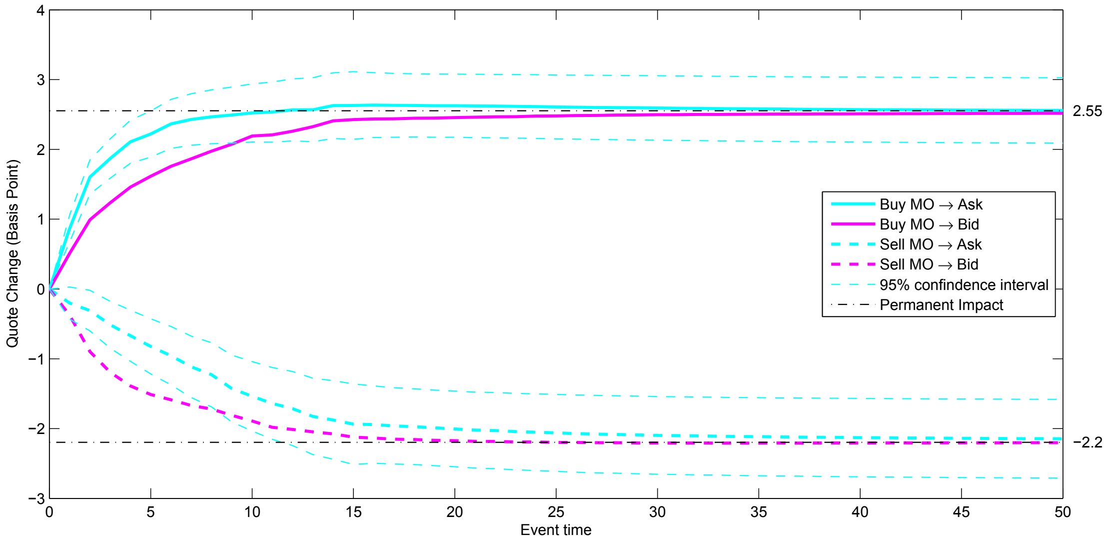


Figure 1.11: Changes of quotes induced by buy/sell market orders (buyer-/seller-initiated trades) with a size equal to half of the depth on their corresponding first levels. The marked number on the vertical axes indicates the magnitude of the permanent impact. Trading of Fortis at Euronext, Amsterdam in August, 2008. MO: Market order.

the posted volume or by cancellations on the bid side. Similarly, liquidity demand on the ask side shifts the ask upward by 1.8 basis points. Hence, overall we observe an asymmetric re-balancing of quotes and a corresponding re-widening of the spread.

The right plot of Figure 1.10 compares the effects of buy limit orders of different sizes but with same limit price posted inside of the bid-ask spread and thus improving bid quotes again by 5 basis points. We observe quite different impulse response patterns in dependence of the order size. In case of a comparably small order, the posted volume is quickly picked up, shifting back the bid quote. In contrast, large volumes overbidding the prevailing quote cause a long-term upward movement of the bid. Relative to the initial shift of the bid we observe a further approximately 20% price increase. Hence, extraordinary large orders are not likely to be picked up and induce a rather strong buy pressure moving the market upwards. For smaller (though still comparably large) orders, adverse selection and signaling effects seem to counterbalance each other. As a consequence, the bid quote is hardly changed and the long run effect is close to the immediate price improvement.

1.5.3 Market Impact of Trades

Figure 1.11 shows the market impacts induced by buy and sell market orders. We assume that the trade sizes correspond to 50% of the prevailing depth. Consequently, these market orders do not 'walk through' the limit order book and thus ask and bid quotes are unaffected. The quote adjustments shown in Figure 1.11 are subsequent quote responses to trade arrivals. Both bid and ask increase (decrease) sharply after the arrival of a buy (sell) market order. Hence, the arrival of a buy (sell) market order induces aggressive posting on the bid (ask) side resulting in further buy (sell) market orders and limit orders posted inside of the spread. Similar to the findings for limit orders, we find evidence for asymmetric adjustments of the two sides of the market. It turns out that buy market orders shift ask quotes more quickly and strongly than bid quotes. The reverse is true for sell market orders. This result indicates that trades temporarily increase spreads which is in contrast to the effects induced by limit orders. Engle and Patton (2004) report similar findings by analyzing quote data from the NYSE. They show that trades have a positive impact on spreads. Because they directly impose the spread as the underlying cointegration relation of quotes, the price impact of trades on spreads is transitory in their model. Using impulse-response analysis based on a structural VEC model, Escribano and Pascual (2006) also find that spreads permanently widen after the arrival of trades. Note that these effects contradict implications of asymmetric-information-based market microstructure models, such as Glosten and Milgrom (1985) and Easley and O'Hara (1992), where trades should resolve the uncertainty regarding existing information and should result in declining spreads.

The left plot of Figure 1.12 depicts quote reactions induced by an aggressive market order 'walking through' the book (Scenario 4 in Section 1.3.2). It absorbs the best ask level and shifts the best quote to the originally second best level which is assumed to be 10 basis points higher than the previous best ask. Similarly to the effects induced by aggressive limit orders we observe that the initial shift of the best ask is reverted back by approximately 35% inducing a long-run ask increase of 6.5 basis points. Simultaneously, aggressive posting on the bid side shifts bid quotes upward. Hence, the initially widened spread reverts back in an asymmetric way causing more quote movements on the bid side than on the ask side. The responses mirror the corresponding effects induced by aggressive buy limit orders (see Figure 1.10), where the spread is initially narrowed and then asymmetrically re-widened causing also more movements on the bid side than on the ask side.

The right plot of Figure 1.12 compares the market impacts on ask quotes induced by a buy market order in situations of different depth behind the market. It is assumed that the order just absorbs the first ask level and thus induces an instantaneous ask price increase by 5 basis points. In line with the results discussed above, in all three scenarios the initially shifted ask quote is reverted back. However, it turns out that the magnitude of this quote reversion critically depends on the prevailing depth behind the market. In fact, the existence of a huge level-two-depth reverts the ask quote back by approximately 60%. We explain this fact by a strong sell pressure induced by huge sell volume queued on the ask side. Conversely, in case of only little depth prevailing behind the market, the existing sell pressure is weaker causing the incoming buy order to (upward) shift the market more strongly. In the extreme case of a very thin market, we even observe an additional quote increase.

A practical problem faced by many market participants is the fundamental choice

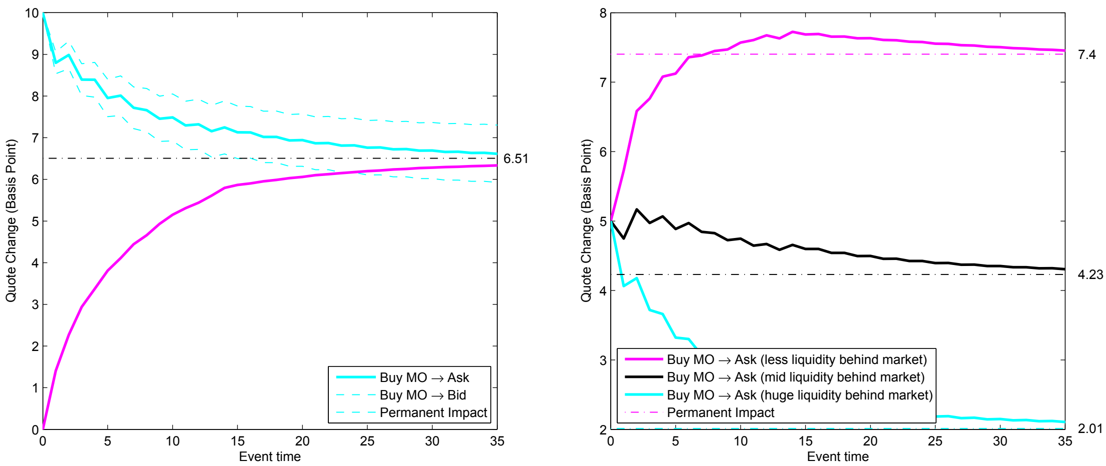


Figure 1.12: Left: Changes of bid and ask quotes induced by an aggressive buy market order with a size exceeding the depth at the best ask by 20%. The second best ask is assumed to be 5 basis points higher than the ask, where the depths behind the market are 1.5 times of the depth at the market. **Right:** Changes of ask quotes induced by an aggressive buy market order with a size equal to the depth at the ask when there is different depth at the second best ask. Case 1: the depth at the second best ask is 10% of that at the ask; Case 2: the depth at the second best ask equals to that at the ask; Case 3: the depth at the second best ask is 500% of that at the ask. The marked number on the vertical axes indicate the magnitude of the permanent impact. Trading of Fortis at Euronext, Amsterdam in August, 2008. MO: Market order.

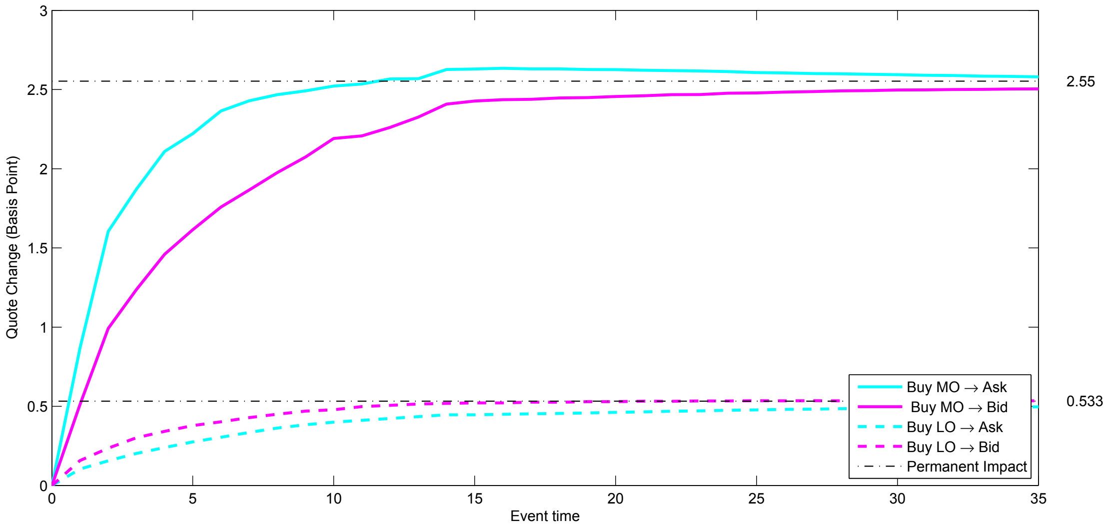


Figure 1.13: Changes of ask and bid quotes induced by a buy market order and a buy limit order of similar size placed at the market. The order size is half of the depth at the best bid. The depths at the best bid and the best ask in the order book are assumed to be equal. Trading of Fortis at Euronext, Amsterdam in August, 2008. LO: limit order; MO: market order.

between posting a market order and a limit order. A direct comparison of the market impacts induced by these two types of orders is shown in Figure 1.13. In both cases, the posted order does not *directly* change quotes. We observe that the resulting long-run effect of trades is significantly greater than that of an equal-size limit order. Actually, the price shift induced by a market order is approximately four times larger than that of a comparable limit order. This finding is consistent with the theoretical prediction by Rosu (2010). Moreover, market orders also cause quicker market reactions. Finally, inferred from the 'gap' between ask and bid curves, market orders change the spread more dramatically than limit orders. Hence, the willingness to cross the bid-ask spread is a stronger signal for private information than that induced by a comparable limit order.

Note that the comparison holds for 'normal' order types placed on the best quote, but not necessarily for more aggressive orders. As discussed above, the long-term effects of aggressive limit orders and market orders critically depend on their (relative) size and the current state of the book. Therefore, an ultimate comparison of market impacts induced by both types of orders under comparable conditions is rather difficult. Nevertheless, our results show that limit orders do have a significant long-term effect and can significantly 'scare' the market.

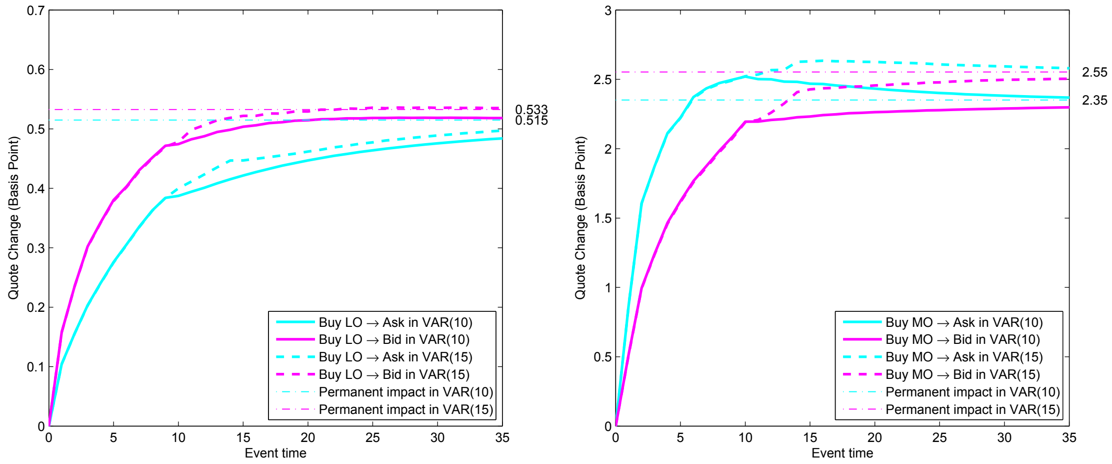


Figure 1.14: Robustness of results. Market impacts of a bid limit order estimated by a VAR(15) and a VAR(25) specification. Trading of Fortis, Euronext Amsterdam in August, 2008.

1.5.4 Robustness of Results

Selecting the appropriate lag order in VAR models is cumbersome in practice when a substantial cross-section of stocks is analyzed over a comparably long period. In order to analyze the sensitivity of our results regarding the choice of the lag order in the VAR model, Figure 1.14 compares the market impacts of a bid limit order and that of a normal buy market order predicted by a VAR(15) model with those induced by a VAR(10) specification using trading of Fortis in August, 2008. It turns out that despite a misspecification of the lag length and remaining serial correlation in the residuals, the impulse response estimates of a VAR(10) are quite close to that of a VAR(15). This is in line with results reported by Jorda (2005) using a VAR(2) to estimate impulse-response functions of an underlying VAR(12) model.

1.5.5 Cross-Sectional Evidence

The complete empirical analysis has been conducted for 29 other stocks traded at Euronext Amsterdam using a VARX(15) specification. The corresponding results are shown in the appendix on the companion web site at `http://amor.cms.hu-berlin.de/~huangrui/project/impact_of_orders/`. It turns out that the results reported in the previous sections are qualitatively stable and representative for a wide cross-section of stocks. Nevertheless, we observe that the *magnitudes* of market impacts vary across the market and seem to be driven by underlying liquidity characteristics. To gain insights into these relationships, we run a simple cross-sectional regression of absolute average market impacts on the average stock-specific trading frequency, trading volume as well as the minimum tick size. I.e.,

$$M_i = \gamma_0 + \gamma_1 N_i + \gamma_2 S_i + \gamma_3 V_i + \varepsilon_i, \tag{1.13}$$

where M_i denotes the absolute permanent impact of stock i induced by a buy/sell order, N_i is the average number of trades per day, S_i represents the normalized tick size, and V_i denotes the normalized trade volume per day. Particularly,

$$S = \frac{\text{tick size} \times 100}{\text{average of closing prices}}, \quad V = \frac{\text{adjusted trading volume per day}}{\text{number of outstanding shares}} \times 100.$$

The scenarios we consider below are similar to those studied in Section 1.3.2. The initial order book for each stock equals its monthly average.

Scenario 'normal limit order' and 'normal market order' : These scenarios are identical to that in Section 1.3.2.

Scenario 'aggressive limit order' : An incoming order of a size which is half to the depth at the corresponding best price is posted inside of the spread and improving the corresponding quote by one tick.

Scenario 'aggressive market order' : An incoming market order with a size equal to the depth at the corresponding best price and thus absorbing the first level in the book.

For every scenario, we consider average market impacts of both buy and sell orders for 30 stocks estimated over two months resulting in 120 observations for each regression.

Table 1.7 reports the corresponding estimation results for two versions of the model: one with included trading volume and one without. The high R^2 values, ranging between 67% and 80%, show that most of the cross-sectional variation of market impact can indeed be explained by the three explanatory variables. It turns out that the trading volume (though its parameter is significant) does not provide much explanatory power. This result indicates that the trading frequency rather than the trade size drives the strength of market responses to limit order arrivals. Furthermore, we observe that the trading frequency has a negative influence on the market impact of limit orders. Hence, in case of a slower trading, a single order conveys more information.

The tick size is positively related to the magnitude of permanent impacts in all scenarios. For aggressive limit orders, this finding is not surprising as the implied price improvement is (relatively) higher for stocks trading on larger tick sizes. Since in these cases, also the spreads between best and second best quotes are higher, the immediate price shift by the arrival of an aggressive market order is larger as well. In the scenarios 'normal limit order' and 'normal market order', a higher tick size and thus an increase of the price discreteness makes it more likely that investors are forced to under-react or over-react in response to incoming information inducing higher deviations between quoted prices and the 'true' underlying efficient price. Our findings show that in these situations, investors rather tend to over-react after the arrival of an order.

Table 1.7
Cross-sectional analysis of market impacts over the market.

We consider market impacts of both buy and sell orders for 30 stocks estimated over August and September, 2008. The regression is $M_i = \gamma_0 + \gamma_1 N_i + \gamma_2 S_i + \gamma_3 V_i + \varepsilon_i$, where M_i denotes the absolute permanent impact of stock i induced by a buy/sell limit order, N_i is the average number of trades per day, S_i represents the normalized tick size, and V_i denotes the normalized trade volume per day. The numbers in brackets denote heteroskedasticity robust $t-$statistics according to White (1980).

Scenario	γ_0	γ_1	γ_2	γ_3	R^2
'normal limit order'	0.0033	−0.0013	0.0418	–	0.69
	(19.02)	(−7.30)	(33.72)		
	0.0027	−0.0015	0.040	0.0011	0.73
	(13.77)	(−11.35)	(34.85)	(6.5)	
'aggressive	0.005	−0.0017	0.096	–	0.79
limit order'	(13.12)	(−5.93)	(23.04)		
	0.004	−0.002	0.095	0.001	0.80
	(11.98)	(−7.92)	(22.80)	(8.34)	
'normal market order'	0.037	−0.016	0.17	−−	0.57
	(21.94)	(−9.26)	(6.33)		
	0.030	−0.018	0.151	0.013	0.67
	(14.59)	(−16.74)	(5.56)	(7.49)	
'aggressive	0.049	−0.020	0.474	−−	0.68
market order'	(19.89)	(−8.24)	(14.46)		
	0.039	−0.023	0.449	0.018	0.75
	(13.03)	(−15.06)	(14.07)	(7.22)	

1.6 Conclusion

In this chapter, we quantify the market impact of incoming limit orders in a limit order book market. Best bid and ask quotes as well as three levels of depth on both sides of the market are modeled based on a cointegrated VAR system. Incoming limit orders are represented in terms of shocks to the system. Limit order characteristics as well as the corresponding state of the book are captured by the specific design of the shock vector. This allows us to distinguish between limit orders of different aggressiveness (reflected by their distance to the market) and different sizes as well as between different states of the book. The market impacts on ask and bid prices are quantified by the estimated impulse response function using appropriate statistical inference.

Employing this framework we analyze the limit order book processes of 30 stocks traded on Euronext Amsterdam over two months in 2008. The model is estimated using the highest possible frequency accounting for all order book changes during continuous trading. Parameter estimates and diagnostics indicate that the proposed model captures the high-frequency order book dynamics quite well.

Based on the empirical analysis we can summarize the following findings: First, we find clear evidence for cointegration relations between ask and bid quotes and corresponding depths. While some cointegration relations are similar to the bid-ask spread, others show that depth has a distinct effect on quote dynamics and on the connection between ask and bid quotes. Second, limit orders do have significant long-term effects on quotes. This is even true for limit orders placed behind the market though these effects decline with the limit order's distance to the market. While incoming limit orders temporarily decrease the spread, market orders induce a temporary widening. Third, the speed of spread convergence as well as the direction of price movements after the arrival of aggressive limit orders undercutting (or overbidding, respectively) best ask and bid prices strongly depends on the incoming limit order's size. While small orders seem to face adverse selection risks and are likely to be picked up quickly, for larger orders information signaling effects seem to dominate pushing the market in the opposite direction. Fourth, the decrease (increase) of spreads after the arrival of an aggressive limit (market) order is reverted back asymmetrically inducing more quote movements on the side where the order has been placed. Fifth, the long-run market impact of aggressive market orders walking through the book rises with the queued depth behind the market. Sixth, the effects are qualitatively remarkably stable over the cross-section of the market. Variations in the magnitudes of market impacts are well explained by the underlying stock-specific trading frequency and minimum tick size.

Our empirical results also show that the proposed framework is useful and appropriate to capture order book dynamics on high frequencies. By modeling quotes *and* several levels of depth, the model implicitly captures also the multivariate dynamics of mid-quotes, returns, spreads, spread changes as well as depth imbalances. In this sense, the suggested high-frequency cointegrated VAR model can serve as a workhorse for various applications in this area.

Chapter 2

Limit Order Properties and Optimal Order Sizes

This chapter is based on Hautsch and Huang (2012a).

2.1 Introduction

Electronic limit order book (LOB) systems are the dominant trading form of most financial markets worldwide, including leading exchanges like NASDAQ, NYSE, BATS and Euronext, various Alternative Trading Systems (ATSs) and Electronic Communication Networks (ECNs). The recent decade witnesses substantial technological progress in trading systems as well as trade recording and an increasing importance of intraday trading. Transparency, low latency, high liquidity and low trading costs attract an increasing number of intraday traders, long-horizon traders as well as institutional investors. Though electronic limit order book trading already exists for many years, further developments in trading systems and structures are ongoing and are faster than ever before. The successive automatization of order management and execution by computer algorithms, the growing importance of smart order routing as well as changes of market structures and trading forms challenge empirical and theoretical market microstructure research.

The objective of this chapter is to provide new empirical evidence on order activities and market dynamics at NASDAQ – the largest electronic market for equities in the U.S. By employing TotalView-ITCH data containing information directly stemming from the NASDAQ data feed, our study sheds some light on recent order arrival rates, execution rates, cancellation rates and the price impact of incoming quotes. Particularly the market impact of a limit order is a key parameter for trading decisions and play a crucial role for (algorithmic) trading strategies. Also theoretical studies, such as, e.g., Harris (1997), Parlour and Seppi (2008), Boulatov and George (2008) or Rosu (2010), predict that the revelation of a trading intention by limit order placements can indeed adversely affect asset prices. Despite of its importance, empirical evidence on the influence of incoming limit orders is still limited. Only very recently, Hautsch and Huang (2012b), Eisler, Bouchaud, and Kockelkoren (2011) and Cont, Kukanov, and Stoikov (2011) analyze the price impact of limit orders and find significant effects. In this study, we employ Hautsch and Huang's (2012b) framework, which extends the approach by Engle and Patton (2004) and provides deeper insights into the market impact of limit orders in recent NASDAQ trading. Of particular interest is whether the magnitudes of price impacts identified in other markets are also found in the extremely liquid NASDAQ market and which limit order sizes can be ultimately posted without significantly moving the market.

TotalView-ITCH data contains all order messages and thus allows us to reconstruct the NASDAQ limit order book in a very precise way, particularly accounting for all high-frequency limit order activities including also so-called fleeting orders. The latter are present for only few seconds and have the purpose of testing for hidden orders placed in the bid-ask spread. A detailed analysis of the NASDAQ order flow in October 2010 provides the following major results: First, the number of limit order submissions is twenty to forty higher than the number of trades. Secondly, limit order sizes are typically small and clustered at round lot sizes of hundred shares. Third, more than 95% of all limit orders are cancelled without getting executed with most of them being cancelled nearly instantaneously (less than one second) after their submission reflecting the proliferation of algorithmic trading at NASDAQ. Fourth, volume-weighted execution times are significantly greater than average execution times indicating that large orders face more execution risk than small ones.

The market impact of limit orders is quantified by modeling ask and bid quotes and several levels of depth in terms of a cointegrated vector-autoregressive (VAR) system which is updated in event time. Short-run and long-run quote reactions are quantified by impulse-response functions. As proposed by Hautsch and Huang (2012b), this framework allows us to estimate the impact of specific limit order activities including limit order submissions, cancellations and executions (corresponding to trades) which are represented as shocks to the system. Our empirical results show that the short-run and long-run quote reaction patterns after the arrival of a limit order are indeed quite similar to those, for example found for Euronext Amsterdam (see Hautsch and Huang, 2012b). Buy (sell) limit orders cause permanent quote increases (decreases) and a temporary decline of the spread. Moreover, we find that the permanent impact of a limit order posted at the best quote is in most cases approximately 25% of that of a trade of similar size. However, this magnitude can be much smaller when hidden orders are placed inside of the spread. As on other liquid markets, only aggressive limit orders posted on the first or second order level induce significant price impacts whereas orders posted with greater distance to the market have virtually no effect.

Finally, using the estimates of market impacts, we suggest a way to compute the optimal size of a limit order given its expected price impact. The implied order size is calculated by inverting the closed form of the permanent impact yielding a function of the current limit order book and the given market impact control level. This provides useful information to control risks in trading strategies.

The remainder of this chapter is organized in the following way. Section 2.2 briefly introduces the market environment and the data. Section 2.3 provides an explorative analysis of the order flow. The econometric framework is reviewed in Section 2.4. Section 2.5 gives empirical evidence of short-run and long-run quote reactions on order activities. In Section 2.6, we propose a method to compute the optimal order size subject to its position in the book and the expected market impact. Finally, Section 2.7 concludes.

2.2 Market Environment and Data

The NASDAQ stock market is the largest electronic stock market (in terms of trading volume) in the world. In 2006, its traditional market center, Brut and INET electronic communication networks (ECNs) are integrated into a single system. This system offers a single execution algorithm based on price and time order precedence for both market makers and participants of ECNs. During the continuous trading period between 9:30 and 16:00 EST, limit orders are submitted to a centralized computer system where they are matched to prevailing limit or hidden orders on the opposite side. If there is no match or the standing volume in the system is insufficient to fully execute the incoming order, the remaining order volume is placed in the order book. NASDAQ supports various order types like pure market orders (immediate order execution without a price limit), stop orders (automatic issuing of limit orders or market orders when a given price is reached), immediate-or-cancel (IOC) orders, reserve orders and non-display orders, among others.

In this study, we use TotalView-ITCH data containing rich information on order activities. The database includes limit order submissions, cancellations, executions and hidden order executions for every trading day since 7:00am EST when the system starts accepting incoming limit orders. The system is initialized by an empty order book where all overnight limit orders are resubmitted automatically at the beginning of each day. Therefore, we can exactly reconstruct the order book at any time by aggregating the existing visible limit orders according to their limit prices. Furthermore, NASDAQ TotalView, surpassing NASDAQ Level 2, is the standard NASDAQ data feed for displaying the full order book depth for market participants. Hence, the reconstructed order book exactly represents historical real-time-disseminated order book states. Trades are identified via the records of limit orders and hidden order executions. Since the trading direction of limit orders and hidden orders is recorded, we can exactly identify whether a trade is buyer-initiated or seller-initiated. Finally, TotalView-ITCH data records a unique identification of any limit order which allows to track the order and to compute, for instance, its life-time.[1]

Note that a market order, especially when its order size is comparably large, is likely to be filled by several pending limit orders. This results in multiple limit order executions corresponding to a sequence of same-type (sub-)trades within a short time

[1] The limit order book reconstruction and limit order tracking is performed by the software "LOBSTER" (see Huang and Polak, 2011), which can be freely accessed at `http://lobster.wiwi.hu-berlin.de`.

Table 2.1
Summary Statistics of Selected Stocks
The variables are calculated for each stock using data from the Center for Research in Security Prices (CRSP) database. The sample period covers October, 2010, including 21 trading days. MktCap is the market capitalization by October 1, 2010. AvgVol denotes the average daily volume (in thousand shares). MedTurn is the median daily turnover. AvgPrc denotes the average daily closing price. StdRet gives the standard deviation of daily returns.

	MktCap (billion \$)	AvgVol (1000 shrs)	MedTurn (in %)	AvgPrc (in \$)	StdRet (in %)
GOOG	130.13	4059.6	1.24	575.94	2.56
ADBE	13.66	15132	2.04	27.512	3.11
VRTX	7.02	1909.7	0.90	35.947	1.64
WFMI	6.38	3018.5	1.54	37.636	1.80
WCRX	5.66	2941.1	0.94	23.535	2.27
DISH	3.93	2486.5	1.12	19.412	1.38
UTHR	3.18	747.95	0.97	55.884	1.89
LKQX	2.99	430.38	0.26	21.465	1.21
PTEN	2.63	5445.5	3.26	18.536	2.35
STRA	2.37	514.49	2.55	145.42	4.24

interval. We identify transactions as sub-trades if they occur in less than half a second after the previous trade and have the same initiation types. All corresponding sub-trades are consolidated to a single trade representing a market order. Furthermore, to avoid erratic effects during the market opening and closure, our sample period covers only the continuous trading periods between 9:45 and 15:45.

We select ten assets out of the 200 biggest stocks listed at NASDAQ according to their market capitalization in Oct 1, 2010. To obtain a representative cross-section, we first divide the 200 stocks into twenty blocks, and then randomly select one stock from each category. Table 2.1 summarizes fundamental characteristics of these stocks extracted from the Center for Research in Security Prices (CRSP) database.

2.3 Major Order Flow and Order Book Characteristics

Electronic limit order book markets are characterized by high transparency and low latency. They enable most market participants having a view on the current state of the market via real-time updated order books. Traders' instructions are transmitted to the trading platform and executed with extremely short time delays (usually only a few milliseconds).[2] As a consequence, sophisticated trading strategies minimizing trading

[2] There are indeed numerous brokers providing their clients direct market access (DMA).

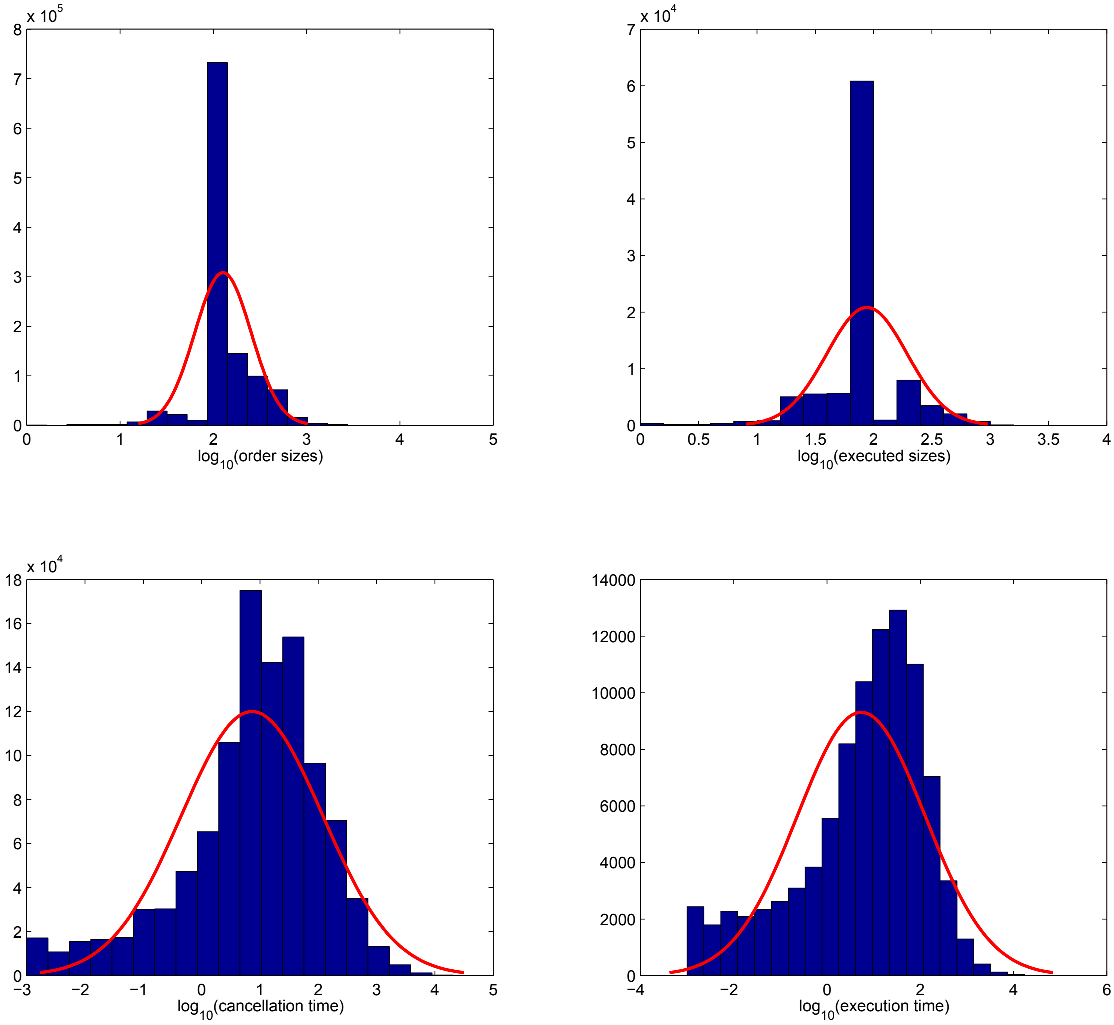


Figure 2.1: Histogram of order sizes, execution sizes, cancellation times and execution times of limit orders. The red line denotes kernel density estimates. Zero cancellation times and execution times are discarded. Trading of WCRX on NASDAQ in October, 2010.

costs and exploiting high-frequency price movements are performed using computer algorithms. Triggered by technological advances, systematic trading is highly sophisticated nowadays. For instance, in order to make profit from high-frequency price fluctuations and liquidity rebates, many high-frequency trading algorithms post a huge number of limit orders which are again canceled almost immediately if not getting executed. As a consequence, enormous limit order activities on extremely high frequencies are observed.

Table 2.2 summarizes the limit order activities of selected stocks and Figure 2.1 shows

Table 2.2

Limit order activities at NASDAQ

Calculated for each stock using TotalView-ITCH messages. The sample period covers October, 2010, including 21 trading days. NumLO is the average daily number of standing limit orders. AvgSZ denotes the average size of limit orders. NumALO is the average daily number of limit orders placed inside the spread ("aggressive" limit orders). NumALO (in %) gives the percentage of aggressive limit orders. NumExe is the number of limit orders getting (possibly partially) executed. MedETim denotes the median execution time of limit orders. VWETim is the volume-weighted execution time. NumCanc (in %) is the percentage of limit orders that are cancelled without (partial) execution. MedCTim denotes the median cancellation time. NumACan (in %) is the cancellation rate of aggressive limit orders. MedACTim gives the median cancellation time of aggressive limit orders.

	NumLO ($\times 10^3$)	AvgSZ (100 shrs)	NumALO ($\times 10^3$)	NumALO (in %)	NumExe ($\times 10^3$)	MedETim (sec.)	VWETim (sec.)	NumCanc (in %)	MedCTim (sec.)	NumACan (in %)	MedACTim (sec.)
GOOG	220.55	1.28	23.50	10.65	5.45	2.77	118.79	97.52	0.42	89.35	0.011
ADBE	206.05	2.48	2.38	1.15	15.28	3.07	107.68	92.57	4.38	50.87	0.351
VRTX	51.59	1.26	3.11	6.03	3.18	6.82	65.67	93.82	8.12	72.49	0.192
WFMI	109.46	1.53	8.06	7.36	5.19	5.92	87.04	95.25	5.88	86.43	0
WCRX	54.50	1.65	1.78	3.27	3.84	10.19	83.87	92.93	10	57.23	0.873
DISH	71.42	1.69	0.91	1.27	3.88	14.35	104.25	94.56	5.55	52.90	0.353
UTHR	27.44	1.20	3.31	12.06	1.22	6.04	52.35	95.54	9.87	81.40	0.352
LKQX	22.92	1.44	1.76	7.68	0.77	12.97	71.15	96.62	14.76	84.37	2.096
PTEN	91.57	1.98	1.81	1.98	7.66	5.71	77.56	91.62	4.84	53.80	0.545
STRA	13.05	1.12	4.02	30.83	0.57	4.89	89.11	95.57	5.17	90.96	1.502

the histograms of constructed variables for one illustrating stock, Warner Chilcott plc (ticker symbol WCRX).[3] The following main findings can be summarized:

(i) Market participants submit a huge number of limit orders with small sizes. The average limit order size is approximately 156 shares. The up-left plot in Figure 2.1 shows that a large proportion of limit orders have a size of 100 shares, corresponding to a round lot on NASDAQ.

(ii) Most of the limit orders are posted at or behind the market. We observe that only approximately 8.2% of the limit orders are placed within the spread and thus update the best quotes.

(iii) Only few limit orders are executed. The median execution time across the ten stocks is approximately 7 seconds. However, the volume-weighted average execution time is substantially greater than its median, reflecting the fact that large limit orders face significantly higher execution risk than small orders.

(iv) More than 95% of limit orders are cancelled without (partial) execution. The median cancellation time of aggressive limit orders placed inside of the spread is less than one second. Hasbrouck and Saar (2009) argue that such a high cancellation rate of limit orders at NASDAQ mainly results from traders 'pinging' for hidden liquidity in the market.

Note that the quickly canceled limit orders change the order book but reverse it back immediately. This nearly instantaneous change is virtually unobservable for humans but can be captured only by trading algorithms run by high-speed computers connecting to exchanges with very low latency.[4] Though such limit order activities do not generally provide any liquidity to the market, they are indispensable for analyzing order book dynamics. Table 2.3 gives summary statistics of market order activities. The number of market orders is substantially smaller than the number of (incoming) limit orders. Interestingly, most market orders are filled by standing limit or hidden orders pending at prices better than or equal to the best quote. Hence, we hardly find market orders walking through the order book.

Table 2.4 gives descriptive statistics of the order book data used in this chapter. We observe significantly more order book updates in the first three order levels than transactions. Moreover, on average, second level market depth is higher than the first level depth while it is lower than the depth on the third level.

[3] The corresponding histograms for other nine stocks are provided on the companion website `http://amor.cms.hu-berlin.de/huangrui/project/order_impact_nasdaq/`. They confirm that our findings are quite consistent across the market.

[4] As a matter of fact, ITCH-Totalview has already reserved time stamps in nano-second precision in order to potentially increase the time resolution in the near future.

Table 2.3
Market order activities
Calculated for each stock using TotalView-ITCH messages. The sample period covers October, 2010, including 21 trading days. NumMO is the average daily number of market orders. AvgSZ denotes the average size of market orders. NumIS (in %) is the percentage of market orders completley filled by hidden orders placed in the spread. NumL1 (in %) is the percentage of market orders filled at the best displayed quote. NumL2 (in %) is the percentage of market orders walking through the book up to the second level. NumL3 (in %) is the percentage of market orders walking through the book up to (or deeper than) the third level.

	NumMO	AvgSZ (100 shrs)	NumIS (in %)	NumL1 (in %)	NumL2 (in %)	NumL3 (in %)
GOOG	6226.4	1.66	43.3	53.8	2.10	0.42
ADBE	4169.1	6.93	4.8	94.9	0.24	0.01
VRTX	1730.0	2.68	13.0	86.4	0.43	0.02
WFMI	2285.7	3.27	7.2	92.2	0.38	0.04
WCRX	1977.0	3.15	8.5	91.0	0.44	0.01
DISH	1339.1	4.45	4.6	95.2	0.15	0
UTHR	857.1	2.16	26.3	72.9	0.67	0
LKQX	469.8	2.23	15.9	83.7	0.29	0.02
PTEN	2647.2	5.36	4.9	95.0	0.07	0
STRA	657.9	1.51	39.8	58.6	1.11	0.12

2.4 An Econometric Model for the Market Impact of Limit Orders

To estimate the market impact of limit orders, we apply the framework proposed by Hautsch and Huang (2012b). The major idea is to model the limit order book in terms of a cointegrated VAR model for quotes and order book depth and to back out the price impact of specific types of limit orders based on impulse response functions.

2.4.1 A Cointegrated VAR Model for the Limit Order Book

Denote t as a (business) time index, indicating all order book activities, i.e., incoming limit or market orders as well as limit order cancellations. Furthermore, p_t^a and p_t^b denote the best log ask and bid quotes instantaneously after the t-th order activity and $v_t^{a,j}$ and $v_t^{b,j}$, $j = 1, \ldots, k$, define the log depth on the j-th best observed quote level on the ask and bid side, respectively. Moreover, to capture dynamic interactions between limit order and market order activities, we define two dummy variables, BUY_t and $SELL_t$, indicating the occurrence of buy and sell trades. Then, the resulting $(4 + 2 \times k)$-dimensional vector

Table 2.4

Summary of order books

The variables are calculated for each stock using reconstructed order book data. The sample period covers October, 2010, including 21 trading days. AvgObs($\times 10^3$) is the average number of observations per day. AvgTrd is the average number of daily trades. AvgAsk is the average of the best ask quote in order books. AvgBid is the average of the best bid quote. AvgSpr (in \$) is the average dollar spread in cents. AvgSpr (in %) is the average relative spread. L1 – L3 denotes the average pending volume on the best quote up to the third best quote.

	AvgObs	AvgTrd	AvgAsk	AvgBid	AvgSpr	AvgSpr	depth on ask (100 shrs)			depth on bid (100 shrs)		
	($\times 10^3$)	($\times 10^3$)	(in \$)	(in \$)	(in cents)	(in %)	L1	L2	L3	L1	L2	L3
GOOG	96.97	3.66	572.45	572.17	28.74	0.051	2.05	1.70	1.56	2.01	1.70	1.58
ADBE	139.52	3.95	27.336	27.325	1.51	0.055	32.87	48.34	63.86	28.86	45.29	61.51
VRTX	39.79	1.51	35.913	35.895	2.25	0.063	4.48	5.19	7.764	4.12	5.13	7.80
WFMI	76.28	2.13	37.59	37.576	1.85	0.049	5.82	8.76	13.61	5.60	8.35	12.56
WCRX	43.32	1.81	23.57	23.558	1.81	0.076	9.22	11.71	15.95	8.31	10.95	15.40
DISH	62.70	1.26	19.396	19.385	1.52	0.078	15.36	18.85	27.88	16.18	18.68	26.54
PTEN	80.33	6.44	18.677	18.666	6.39	0.114	17.02	23.67	28.71	16.65	22.18	27.24
LKQX	19.17	4.01	21.464	21.441	2.70	0.126	3.39	3.96	5.02	3.36	4.18	5.44
UTHR	18.92	2.51	56.083	56.026	1.50	0.080	2.20	2.13	2.40	2.06	2.01	2.18
STRA	16.26	0.40	145.25	144.77	53.33	0.364	1.75	2.09	2.85	1.50	1.68	2.05

of endogenous variables is given by

$$y_t := [p_t^a, p_t^b, v_t^{a,1}, \ldots, v_t^{a,k}, v_1^{b,1}, \ldots, v_t^{b,k}, BUY_t, SELL_t]'. \tag{2.1}$$

The quote levels associated with $v_t^{a,j}$ and $v_t^{b,j}$ are not observed on a *fixed* grid at and behind the best quotes. Consequently, their price distance to p_t^a and p_t^b is not necessarily exactly $j-1$ ticks but might be higher if there are no limit orders on all possible intermediate price levels behind the market.

Note that market depth enters the vector y_t in levels and thus is treated as a possibly non-stationary variable. Since market depth is highly persistent and (on very high frequencies) reveals features of a near-unit-root process, Hautsch and Huang (2012b) recommend treating this variable as being possibly non-stationary. This guarantees consistency of estimates, even if market depth is truly stationary.

Following Hautsch and Huang (2012b) we model the process in terms of a restricted cointegrated VAR model of the order p (VAR(p)) with the Vector Error Correction (VEC) form for $\Delta y_t := y_t - y_{t-1}$,

$$\Delta y_t = \mu + \alpha\beta' y_{t-1} + \sum_{i=1}^{p-1} \gamma_i \Delta y_{t-i} + u_t, \tag{2.2}$$

where u_t is white noise with covariance matrix σ_u, μ is a constant, γ_i with $i = 1, \ldots, p-1$, is a $k \times k$ parameter matrix, and α and β denote the $k \times r$ loading and cointegrating matrices with $r < k$. By treating the trading indicators BUY_t and $SELL_t$ as stationary variables, the two first columns of β are restricted to $\beta_1 = [0, \ldots, 0, 1, 0]'$ and $\beta_2 = [0, \ldots, 0, 0, 1]'$.

The corresponding reduced VAR representation in levels of y_t is given by

$$y_t = \mu + \sum_{i=1}^{p} a_i y_{t-i} + u_t, \tag{2.3}$$

where $a_1 := I_k + \alpha\beta' + \gamma_1$ with I_k denoting a $k \times k$ identity matrix, $a_i := \gamma_i - \gamma_{i-1}$ with $1 < i < p$ and $a_p := -\gamma_{p-1}$. As illustrated by Hautsch and Huang (2012b), the model (2.2) can be estimated by full information maximum likelihood (FIML) according to Johansen (1991) and Johansen and Juselius (1990).

Table 2.5 shows the estimated cointegrating vectors for a representative trading day, where we omit the two known cointegrating vectors associated with the (stationary) trading indicators and all corresponding elements in the remaining cointegration vectors. The resulting vectors are ordered according to their corresponding eigenvalues reflecting their likelihood contributions.

We observe that the first five and the last cointegration relations are mostly linear combinations of spreads and depths. Specifically, the first one is quite similar to a linear combination mimicking the bid-ask spread. The most interesting relationship is implied by the vector $\hat{\beta}_8$, revealing relatively large (and different) coefficients associated with the depth variables. This indicates that depth has a significant impact on the long-term

Table 2.5
Representative estimates of cointegrating vectors
The vectors are sorted according to their corresponding eigenvalues in Johansen's ML approach. Overall there are nine cointegrating vectors. Two of them are known, i.e., $\beta_1 = [0, \ldots, 0, 1, 0]$ and $\beta_2 = [0, \ldots, 0, 0, 1]$, representing the stationary trading indicators. Accordingly, the elements corresponding to BUY and $SELL$, in $\widehat{\beta}_3$ to $\widehat{\beta}_9$ are set to zero and are omitted as well. Trading of WRCX at NASDAQ on October 1, 2010.

Variable	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$	$\hat{\beta}_9$
p^a	1.00	-1.00	1.00	1.00	0.99	-0.64	-0.97
p^b	-0.98	0.99	-0.99	-0.99	-1.00	1.00	1.00
$v^{a,1}$	0.00	-0.25	-0.01	-0.12	-0.03	-0.03	0.02
$v^{a,2}$	-0.00	0.26	0.06	-0.11	0.00	-0.34	0.03
$v^{a,3}$	0.00	-0.18	-0.04	0.19	0.09	-0.54	0.02
$v^{b,1}$	0.01	0.16	-0.03	0.02	-0.06	0.02	0.02
$v^{b,2}$	-0.01	-0.17	0.05	0.15	-0.07	0.37	0.01
$v^{b,3}$	0.01	0.11	-0.02	-0.10	0.13	0.73	-0.00

relationship between quotes. Intuitively, the connection between ask and bid quotes becomes weaker (and thus deviates from the spread) if the depth is less balanced between both sides of the market. Hence, depth has a significant impact on quote dynamics and should be explicitly taken into account in a model for quotes. These results strongly confirm corresponding findings by Hautsch and Huang (2012b) for trading at Euronext.

Finally, note that model (2.3) can be further rotated in order to represent dynamics in spreads, relative spread changes, midquotes, midquote returns as well as (ask-bid) depth imbalances. Hence, the model is sufficiently flexible to capture the high-frequency dynamics of all relevant trading variables. In this sense, the approach complements dynamic models for order book curves such as proposed by Härdle, Hautsch, and Mihoci (2009) and Russell and Kim (2010).

2.4.2 Estimating Market Impact

The market impact of limit orders can be backed out by representing an incoming order as a shock to the dynamic order book system as specified in equation (2.3). Whenever an order enters the market, it (i) will change the depth in the book, (ii) may change the best quotes depending on which position in the queue it is placed, and (iii) will change the trading indicator dummy in case of a market order. Consequently, the direct effects of a limit order can be represented in terms of a 'shock' vector $\delta_y := [\delta_p', \delta_v', \delta_d']'$, where δ_p denotes a 2×1 vector containing shocks in quotes, δ_v is a $2k \times 1$ vector representing shocks in depths, and δ_d denotes a 2×1 vector containing changes of the trading indicator

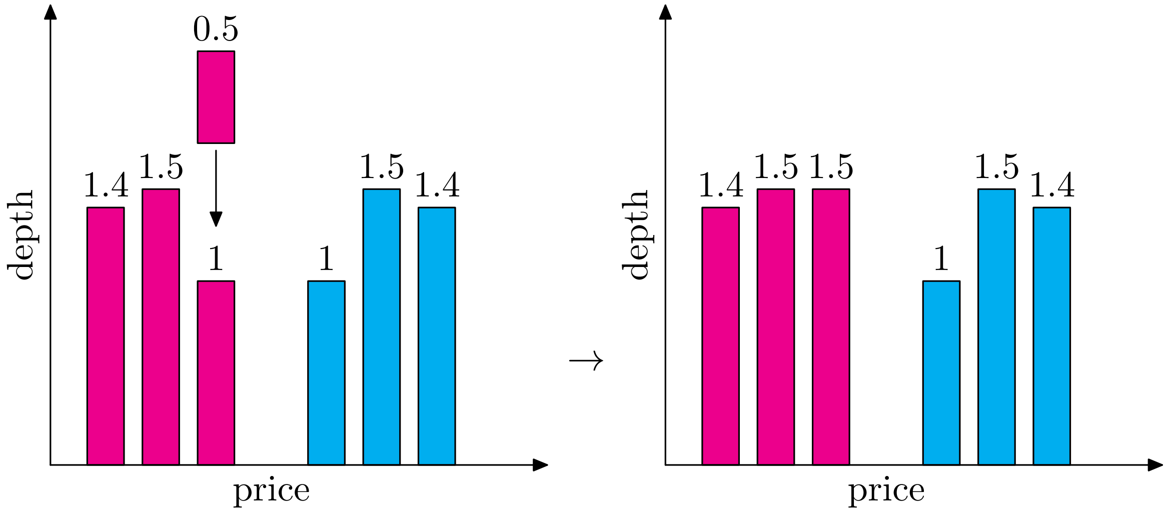


Figure 2.2 (Scenario 1a (normal limit order)): An incoming buy limit order with price 1000 and size 0.5. It affects only the depth at the best bid without changing the prevailing quotes or resulting in a trade. *Figure from Hautsch and Huang (2012b).*

dummy.

Following Hautsch and Huang (2012b), we design the impulse response vectors associated with four scenarios commonly faced by market participants. As graphically illustrated by Figures 2.2 to 2.4, a three-level order book is initialized at the best ask $p_t^a = 1002$, best bid $p_t^b = 1000$, second best ask 1003, second best bid 999, and depth levels on the bid side $v_t^{b,1} = 1$, $v_t^{b,2} = 1.5$, $v_t^{b,3} = v_t^{b,4} = 1.4$. The following scenarios are considered:[5]

Scenario 1a (normal limit order): arrival of a buy limit order with price 1000 and size 0.5 to be placed *at the market*. As shown in Figure 2.2, this order will be consolidated at the best bid without changing the prevailing quotes. Because the initial depth on the first level is 1.0, the change of the log depth is $\ln(1.5) \approx 0.4$. Correspondingly, the shock vectors are given by $\delta_v = [0, 0, 0, 0.4, 0, 0]'$, $\delta_p = \delta_d = [0, 0]'$.

Scenario 1b (passive limit order): arrival of a buy limit order with price 999 and size 0.5 to be posted *behind the market*. As in the scenario above, it does not change the prevailing quotes and only affects the depth at the second best bid. We have $\delta_v = [0, 0, 0, 0, \ln(2) - \ln(1.5) \approx 0.29, 0]'$, $\delta_p = \delta_d = [0, 0]'$.

Scenario 2 (aggressive limit order): arrival of a buy limit order with price 1001 and size 0.5 to be posted inside of the current spread. Figure 2.3 shows that it improves

[5]For sake of brevity, the scenarios are only characterized for buy orders. Sell orders are analyzed accordingly.

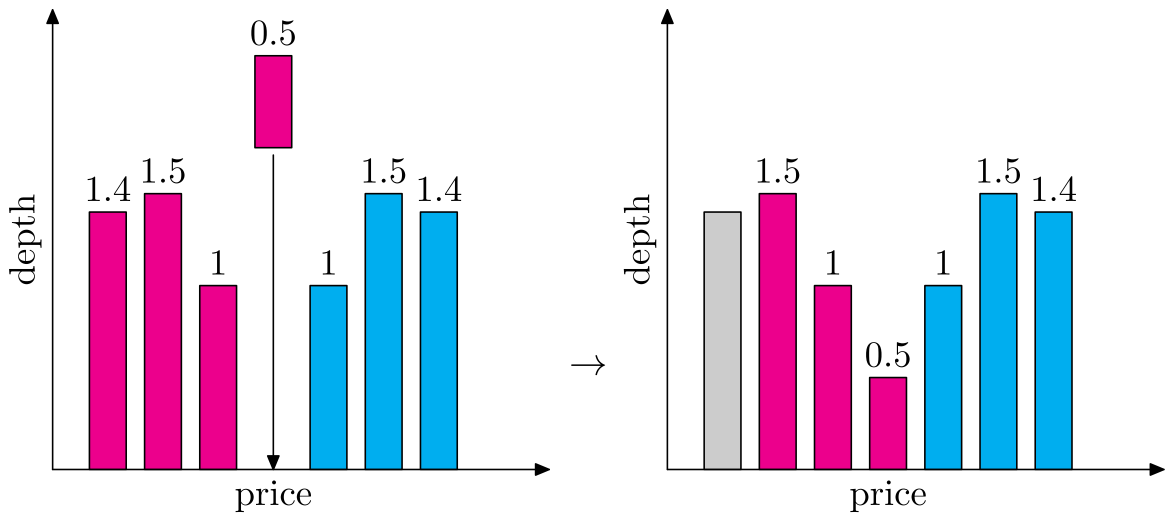


Figure 2.3 (Scenario 2 (aggressive limit order)): An incoming buy limit order with price 1001 and size 0.5 improving the best bid and changing all depth levels on the bid side of the order book. *Figure from Hautsch and Huang (2012b).*

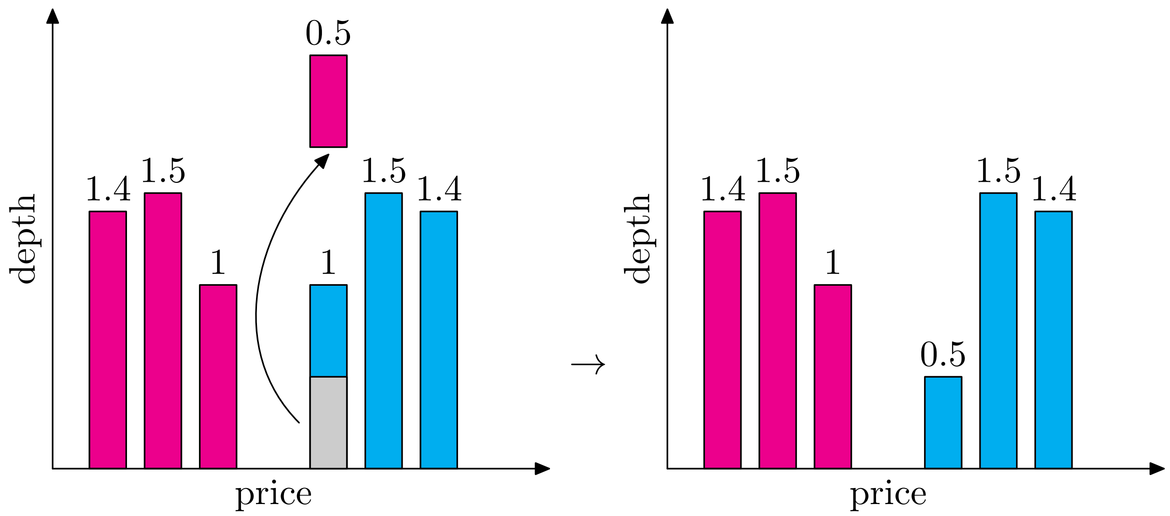


Figure 2.4 (Scenario 3 (normal market order)): An incoming buy market order with price 1002 and size 0.5 which results in a buyer-initiated (buy) trade. *Figure from Hautsch and Huang (2012b).*

Table 2.6
Shock vectors implied by the underlying four scenarios

Initial order book: best ask $p_t^a = 1002$, best bid $p_t^b = 1000$, second best ask = 1003, second best bid = 999. Volumes on the ask/bid side: $v_t^{a/b,1} = 1$ at the best bid, $v_t^{a/b,2} = 1.5$ at the second best bid, and $v_t^{a/b,3} = v_t^{a/b,4} = 1.4$ at the third and fourth best bids, respectively. Notation: δ_v denotes changes in market depths; δ_p denotes changes of the best bid and best ask; δ_d denotes changes of the trading indicator variables. *Figure from Hautsch and Huang (2012b)*

Scenario	limit order	shock vectors		
	(dir,price,size)	δ_v'	δ_p'	δ_d'
'normal limit order'	(Bid,1000, 0.5)	$[0, 0, 0, 0.4, 0, 0]$	$[0, 0]$	$[0, 0]$
'passive limit order'	(Bid,999, 0.5)	$[0, 0, 0, 0, 0.29, 0]$	$[0, 0]$	$[0, 0]$
'aggressive limit order'	(Bid,1001, 0.5)	$[0, 0, 0, -0.69, -0.4, 0.07]$	$[0, 0.001]$	$[0, 0]$
'normal market order'	(Bid,1002, 0.5)	$[-0.69, 0, 0, 0, 0, 0]$	$[0, 0]$	$[1, 0]$

the best bid by 0.1% and accordingly shifts all depth levels on the bid side. The resulting shock vector is given by $\delta_v = [0, 0, 0, (\ln(0.5) \approx -0.69), (\ln(1/1.5) \approx -0.4), (\ln(1.5/1.4) \approx 0.07)]'$, $\delta_p = [0, 0.001]'$ and $\delta_d = [0, 0]'$.

Scenario 3 (normal market order): arrival of a buy order with price 1002 and size 0.5. This order will be immediately executed against standing limit orders at the best ask quote. Because it absorbs liquidity from the book, it shocks the corresponding depth levels negatively. Figure 2.4 depicts the corresponding changes of the order book as represented by $\delta_v = [\ln(0.5) \approx -0.69, 0, 0, 0, 0, 0]'$, $\delta_p = [0, 0]'$ and $\delta_d = [1, 0]'$.

Table 2.6 summarizes the shock vectors implied by the illustrating scenarios.

The market reactions induced by incoming limit orders are captured by the impulse response function,

$$f(h; \delta_y) = \mathrm{E}[y_{t+h}|y_t + \delta_y, y_{t-1}, \cdots] - \mathrm{E}[y_{t+h}|y_t, y_{t-1}, \cdots], \tag{2.4}$$

where the shock on quotes, depths and trading indicators is denoted by $\delta_y := [\delta_p', \delta_v', \delta_d']'$ and h is the number of periods (measured in 'order event time').

Note that the impulse responses need not to be orthogonalized as contemporaneous relationships between quotes and depths are captured by construction of the shock vector. Moreover, our data is based on the arrival time of orders avoiding time aggregation as another source of mutual dependence in high-frequency order book data. The impulse-response function according to equation (2.4) can be written as

$$f(h; \delta_y) = J\mathbf{A}^h J'\delta_y, \tag{2.5}$$

where

$$\mathbf{A} := \underbrace{\begin{bmatrix} A_1 & \cdots & A_{p-1} & A_p \\ I_K & & 0 & 0 \\ & \ddots & \vdots & \vdots \\ 0 & \cdots & I_K & 0 \end{bmatrix}}_{Kp\times Kp}.$$

Given the consistent estimator $\widehat{a}$ for $a := \text{vec}(A_1, \ldots, A_p)$ in (2.3) we have

$$\sqrt{T}(\widehat{a} - a) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widehat{a}}).$$

Lütkepohl (1990) shows that the asymptotic distribution of the impulse-response function is given by

$$\sqrt{T}(\hat{f} - f) \xrightarrow{d} \mathcal{N}(0, G_h \Sigma_{\widehat{a}} G_h'), \tag{2.6}$$

where $G_h := \partial\,\text{vec}(f)/\partial\,\text{vec}(A_1, \cdots, A_p)'$. This expression can be explicitly written as

$$G_h = \sum_{i=0}^{h-1} \left(\delta_y' J(\mathbf{A}')^{h-1-i} \otimes J\mathbf{A}^i J'\right). \tag{2.7}$$

The permanent impact of limit order can be deduced from Ganger's representation of the cointegrated VAR as

$$\bar{f}(\delta_y) := \lim_{h\to\infty} f(h; \delta_y) = C\delta_y, \tag{2.8}$$

where

$$C = \beta_\perp \left(\alpha_\perp' \left(I_K - \sum_{i=1}^{p-1} \Gamma_i\right) \beta_\perp\right)^{-1} \alpha_\perp' . \tag{2.9}$$

2.5 Market Impact at NASDAQ

We model the best quotes and market depths up to the third level. Computational burden is reduced by separately estimating the model fore each of the 21 trading days. The market impact is then computed as the monthly average of individual (daily) impulse response. Likewise, confidence intervals are computed based on daily averages. For sake of brevity we refrain from presenting all individual results for the ten stocks. We rather illustrate representative evidence based on Warner Chilcott plc (ticker symbol WCRX) using a cointegrated VAR(10) model. The results for the remaining stocks are provided in a web appendix on `http://amor.cms.hu-berlin.de/~huangrui/project/order_impact_nasdaq/`. Indeed, we find the same pattern as in Hautsch and Huang (2012b).

Figure 2.5 depicts the market impact of buy and sell limit orders posted at the best quotes as shown in Scenario 1 in Section 2.4.2. The impact starts at zero since such a limit order does not *directly* change quotes. As expected, both ask and bid quotes

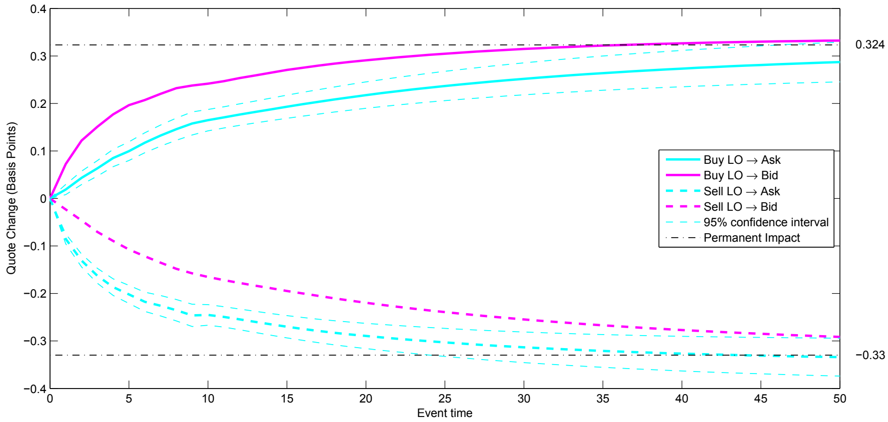


Figure 2.5: Changes of ask and bid quotes induced by buy/sell limit orders placed at the market (level one) with a size equal to the half of the depth on the first level. The marked number on the vertical axes indicates the magnitude of the permanent impact. The blue dotted lines indicate the corresponding 95%-confidence intervals. Trading of WCRX at NASDAQ in October 2010. LO: limit order.

significantly rise (decline) after the arrival of a buy (sell) limit order. In the long-run, both quotes converge to a permanent level at which the information content of the incoming limit order is completely incorporated. We observe that the long-run price change is approximately 0.3 basis points. In the short-run, ask and bid quotes adjust in an asymmetric way where bid (ask) quotes tend to react more quickly than ask (bid) quotes after the arrival of a buy (sell) limit order. This adjustment induces an one-sided and temporary decrease of the bid-ask spread.

To explore the role of the order's position in the book, Figure 2.6 depicts the impact on the bid quote induced by a buy limit order placed at the market (level one) and behind the market (level two and three). We observe that the magnitude and speed of the quote reaction are negatively correlated with the order's distance from the spread. Specifically, for orders posted deeper than the third level in the order book, virtually no market impacts can be identified.

Limit orders placed inside the spread perturb the order book dynamics in a more complex way as show in Scenario 2 in Section 2.4.2. They directly improve the ask or bid resulting in an immediate narrowing of the spread and a shift of one side of the order book. Hence, the system seeks the new equilibrium on a path recovering from an immediate quote change and a simultaneous re-balancing of liquidity. Figure 2.7 shows the reactions of bid and ask quotes induced by an aggressive buy limit order. Given our setting, a buy limit order induces a 4.3 basis point increase of the bid quote (corresponding to approximately one cent). However, the long-run price change is just 2.11 basis points. The immediate quote reversal is induced either by sell trades picking up the volume or by cancellations on the bid side. Likewise the ask quote shifts upward.

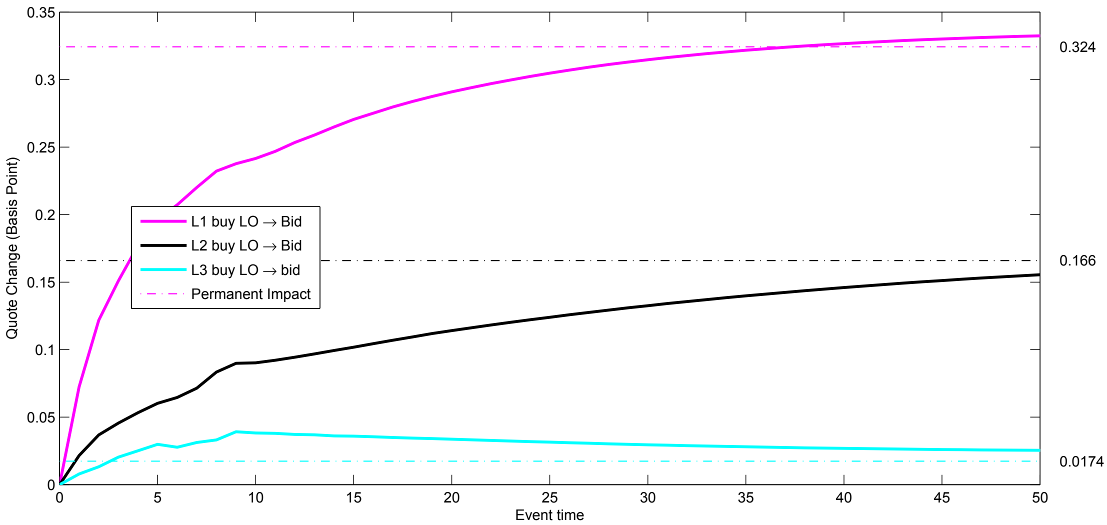


Figure 2.6: Changes of bid quotes induced by buy limit orders placed at the market (level one) and behind the market (level two and three). The order size equals half of that at the best bid. The initial order book equals the corresponding monthly average shown in Table 2.4. The marked number on the vertical axes indicates the magnitude of the permanent impact. Trading of WCRX at NASDAQ in October, 2010. L1: level one. L2: level two. L3: level three. LO: limit order.

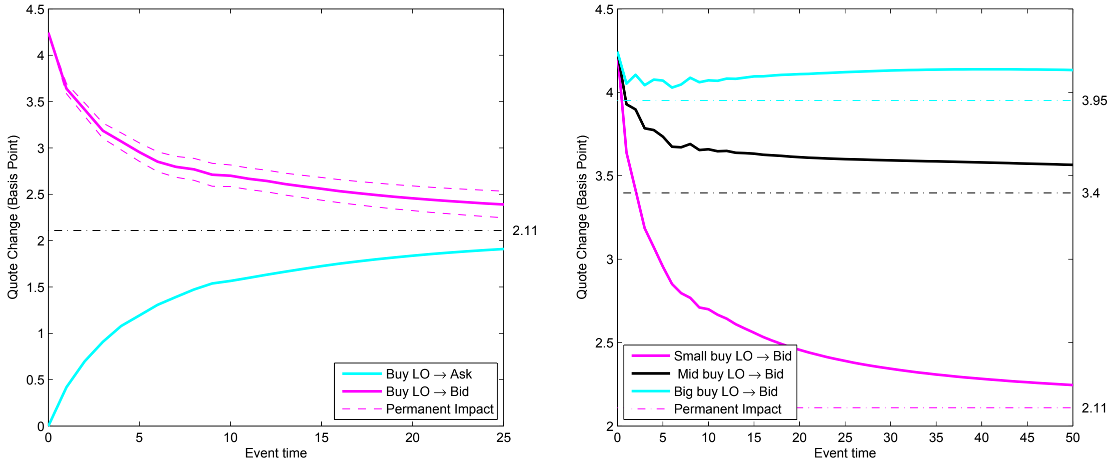


Figure 2.7: Left: Changes of quotes induced by buy limit orders placed inside of the spread with a size equal to the depth at the bid. **Right:** Changes of bid quote induced by buy limit orders placed inside of the spread with different sizes. The initial order book equals the corresponding monthly average shown in Table 2.4. Small size: depth at the bid. Mid size: 7 times of the depth at the bid. Big size: 15 times of the depth at the bid. Trading of WCRX at NASDAQ in October, 2010. LO: limit order.

We hence observe an asymmetric re-balancing of quotes and a corresponding re-widening of the spread.

The right plot of Figure 2.7 shows how the size of incoming aggressive limit orders affects quote reactions. In case of a comparably small order, the posted volume is likely to be quickly picked up or cancelled, shifting back the bid quote. In contrast, large volumes over-bid the prevailing quote causing a significant long-run impact. This confirms findings by Hautsch and Huang (2012b) for Euronext Amsterdam and shows that aggressive limit orders with large order sizes carry information and serve as pricing signals.

Figure 2.8 compares the market impact induced by a buy market order and a similar buy limit order posted at the bid. We observe that both bid and ask quotes sharply increase after the arrival of a buy market order. The permanent shift of quotes induced by a market order is approximately 4 times greater than that by an incoming limit order. This finding supports theoretical predictions by Rosu (2010). Moreover, in case of a market order, the ask reacts more quickly than the bid. Hence, we observe an asymmetric adjustment of the two sides of the market resulting in a temporary widening of spreads.

Since the market impact of limit orders depends not only on the market microstructure but also on the characteristics of the individual stock, an ultimate comparison of estimated market impacts on NASDAQ with those on Euronext (see Hautsch and Huang (2012b)) is rather difficult. Nevertheless, we do find a significant difference when com-

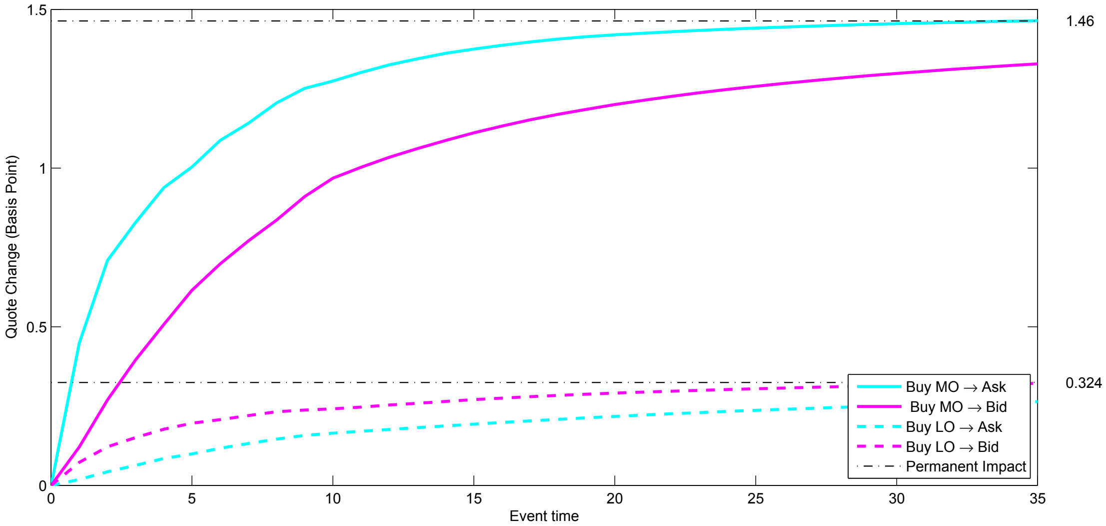


Figure 2.8: Changes of ask and bid quotes induced by a buy market order and a buy limit order of similar size placed at the market. The order size is half of the depth at the best bid. The initial order book equals the corresponding monthly average shown in Table 2.4. Trading of WCRX at NASDAQ in October, 2010. LO: limit order; MO: market order.

paring the market impact of trades to that of limit orders. While on Euronext Hautsch and Huang (2012b) find robust evidence for the market impact of trades trading at best quotes being approximately four times of the market impact of a limit order of similar size, this does not necessarily hold for all stocks at NASDAQ, such as, e.g., GOOG, STRA and UTHR. We explain this finding by the existence of hidden liquidity inside of bid-ask spreads as shown in Table 2.3. When the market participant expects a better price than the best quote to be available inside of the spread, she would naturally interpret a market order placed at the best quotes as being comparably more aggressive as it walks through the (hidden) price levels. As a consequence, the reaction to an incoming market order becomes stronger. Similarly, an incoming limit order is interpreted as being comparably more passive. Consequently, the market impact of limit orders decreases.

2.6 Optimal Order Size

The expected price impact induced by a limit order placement is a key parameter in trading decisions. Therefore, in trading strategies, it might be of particular interest to explicitly control the expected market impact. The estimates of the price impact provided in the previous section can be used to back out the size of an order (given its position in the queue) which is necessary to cause a given expected price impact.

In fact, due to the discreteness of prices, the magnitude of a price impact can be interpreted in a probabilistic context. Given a minimum tick size at equity markets like NASDAQ, a practitioner who prefers not to shift the price with probability ξ must design the order such that the expected price shift, i.e., the magnitude of the impact, is less than $1-\xi$ ticks. This is straightforwardly seen by noticing that when the probability is exactly ξ, the minimum level of the market impact is

$$\begin{aligned} \text{permanent market impact} &= \mathrm{E}[\text{long-run price shift}] \\ &= \xi \times (0 \text{ ticks}) + (1-\xi) \times (1 \text{ ticks}) \\ &= 1 - \xi \quad (\text{ticks}). \end{aligned} \tag{2.10}$$

In the following we shall illustrate how to explicitly compute the optimal order size subject to the given control level ξ. For ease of illustration, consider a bid limit order with size m placed at the second best bid. In our setting based on a 3-level order book, it is represented as a 10-dimensional shock vector with only one non-zero element at the 9-th row according to the order of variables in equation (2.1),

$$\delta_9 = \log\left(\frac{m}{\text{depth at second best bid}} + 1\right).$$

By equation (2.8), the corresponding permanent impact on the bid is given by

$$c_{29} \times \log\left(\frac{m}{\text{depth at second best bid}} + 1\right) \times \left[\frac{\text{bid}}{\text{tick size}}\right], \tag{2.11}$$

where c_{29} is the ninth element in the second row of matrix C in (2.9). Plugging (2.11) into (2.10) and solving for m gives

$$m = \left(\exp\left[\frac{(1-\xi) \times \text{tick size}}{\text{bid} \times c_{29}}\right] - 1\right) \times (\text{depth at second best bid}). \tag{2.12}$$

Figure 2.9 depicts the permanent impact on bid prices against order sizes for the ten selected stocks. Each curve in the sub-plots presents the permanent impact induced by the particular type of bid limit order, i.e., "limit orders placed at the second best bid", "limit orders placed at the best bid" and "limit orders placed inside of the spread". The order book is initialized at its average. For the sake of clarity, we change the unit of impacts (on the $y-$axis) from basis points of bid prices to the number of ticks. Furthermore, the control level ξ is set to 0.9 (corresponding to a permanent market impact of 0.1 ticks) represented by the horizontal dashed line. The intersections S_1 and S_2 correspond to optimal sizes of limit orders placed on the best bid and second best bid, respectively. For instance, for WCRX and subject to the condition that the market impact is less than 0.1 cent, the optimal size for a limit order placed at the best bid is around 600 shares. Likewise, the optimal size for a limit order placed at the second bid is around 1400 shares.

For the stocks GOOG, UTHR and STRA, we observe that the market impact is so large that the intersection S_1 corresponds to an order size of less than 100 shares.

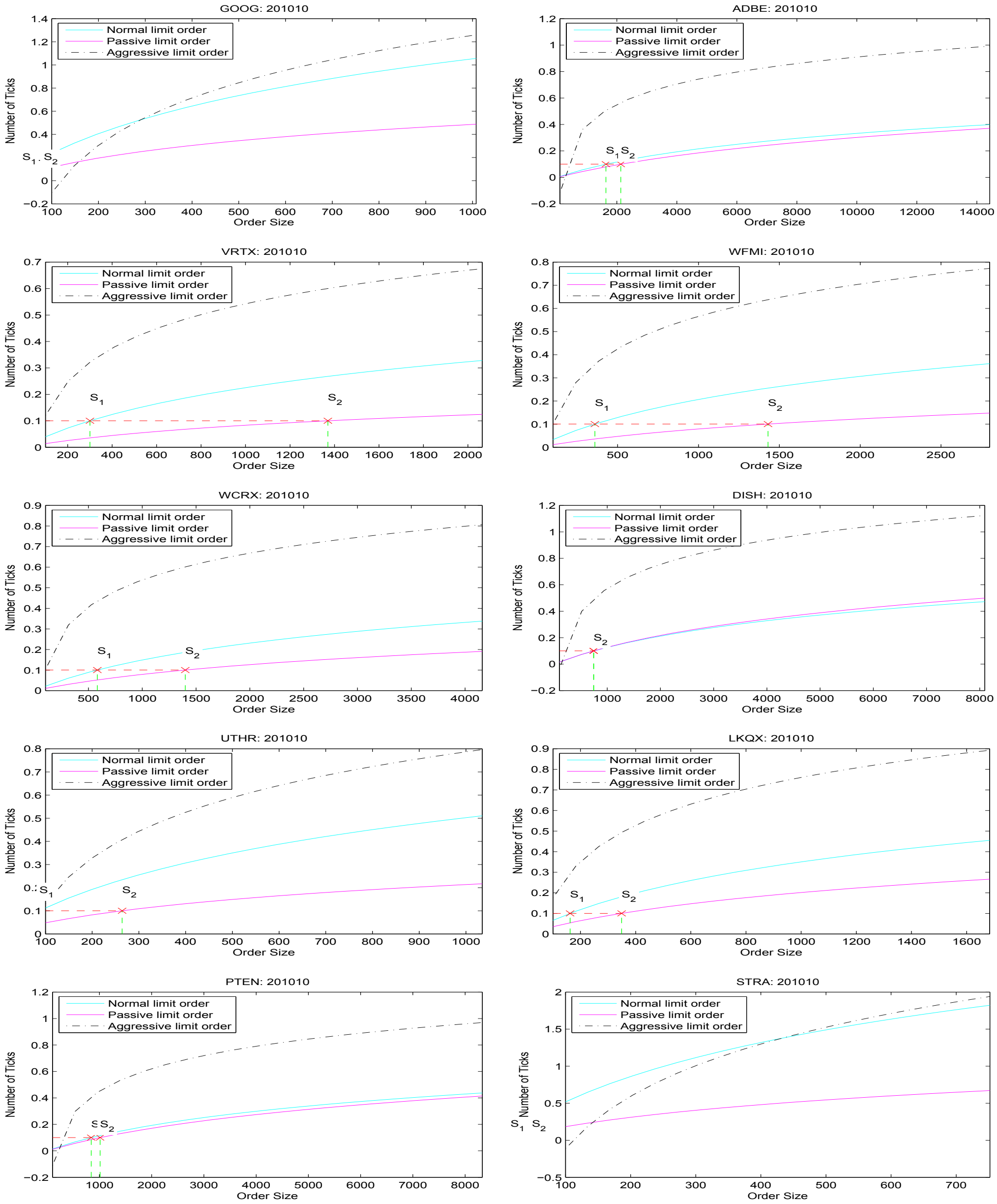


Figure 2.9: Permanent impacts against order sizes. The impacts are induced by bid orders. The initial order book is set to its monthly average. The order sizes at the $x-$axis range from 100 shares to 5 times of the depth at the best bid in the initial order book . The aggressive (in-the-spread) limit orders improve the bid price by 1 cent. The horizontal dashed line presents a subject control level corresponding to a permanent market impact of 0.1 cents. Trading of ten selected stocks at NASDAQ in October, 2010.

We explain this phenomenon by three reasons. First, the depth at the best bid is comparably small. Therefore, a 100-shares-order is a relatively large order given the available liquidity at the market. Second, as shown in Table 2.1, prices of these stocks are relatively high. Consequently, the relative minimum tick size is comparably small implying lower costs of front-running strategies. Hence, the high market impact reflects a high probability of being affected by front-running. Third, the average absolute spread in ticks is large. Consequently, there is more room for other market participants to improve their quotes.

Finally, for some stocks, we observe zero or even negative permanent impacts of small orders placed inside of the spread, as, e.g., GOOG, ADBE, DISH, PTEN and STRA. This is caused by the effect that small limit orders placed inside of the spread are mainly submitted by trading algorithms and tend to be canceled very quickly if not getting executed. In other situations, they might be quickly picked up and trigger other algorithms issuing market orders and/or canceling existing limit orders on their own side.

2.7 Conclusion

In this chapter, we provide new empirical evidence on limit order submissions and market impacts in NASDAQ trading. Employing TotalView-ITCH data, we can summarize the following major findings. First, we observe huge numbers of order submissions per day with order sizes clustering around round lots. Second, most of the limit orders are cancelled before getting executed. Cancellation times are hardly greater than one second. Third, the volume-weighted execution time of limit orders is substantially greater than its median indicating that big limit orders face clearly more execution risk. Finally, we observe that only very few market orders tend to 'walk through the book'.

We find the short-run and long-run price reactions induced by limit order placements to be consistent with those found by Hautsch and Huang (2012b) for data stemming from the Euronext Amsterdam. This implies that these effects are quite stable across markets, despite of differences in market settings. In particular, we find that incoming limit order have significant short-run and long-run effects on ask and bid quotes. Buy (sell) limit orders increase (decrease) both ask and bid quotes while temporarily decreasing bid-ask spreads. Similar but stronger effects are found after arrivals of market orders with temporary increases of bid-ask spreads. For aggressive limit orders posted in the spread we find different effects depending on the order size. While the new quote level caused by a large aggressive order also holds in the long run, this is not true for small orders. Their direct effect on quotes tend to be reversed after a while as the order is picked up. Moreover, it turns out that only limit orders posted up to the second order level have significant market impacts. Orders which are placed even deeper in the book have virtually no effect on the market. Interestingly, we find that small orders placed inside of the spread cause zero or even negative long-run impacts. We explain this finding by the existence of trading algorithms which cancel such orders very quickly if they do not get executed.

Finally, we illustrate how to use the setup to compute optimal sizes of limit orders given a certain intended price impact. This might be helpful to control the risk in trading strategies.

Chapter 3

Identifying and Analyzing Hidden Order Placements

This chapter is based on Hautsch and Huang (2012c).

3.1 Introduction

Since the introduction and the growing dominance of electronic trading during the nineties, equity markets have trended toward higher transparency and more disclosure of trading information. However, displayed limit orders reveal trading intentions and may induce adverse selection effects, picking-off risks and "parasitic trading" (see e.g., Harris, 1997). Consequently, the question of how much transparency should be optimally provided on a market is of ongoing importance. In particular, current developments in equity markets away from full transparency and back toward more opaque market structures made this question again very topical in recent market microstructure research.

In modern trading, traders seek to conceal trading strategies and to avoid adverse price effects by hiding order sizes. Consequently, reserve ("iceberg") orders which require to display only a small fraction of the order size are increasingly popular and can be used on virtually all major exchanges and trading platforms worldwide. An even more extreme form of reducing pre-trade transparency is to trade in form of non-display ("hidden") orders which can be entirely hidden. Such orders do not even reveal the posted limit price and thus act as completely hidden liquidity supply in the order book. While there are a few empirical studies analyzing iceberg orders (see, e.g., Bessembinder, Panayides, and Venkataraman, 2009; Frey and Sandås, 2009), there is no empirical evidence on non-display orders. Due to the possibility to also hide the underlying limit price, they induce trading and order placement effects which are quite different from those caused by iceberg orders and which have not been not well understood yet. For instance, the most interesting aspect behind hidden orders is that they can be placed inside of the bid-ask spread without affecting visible best ask and bid quotes. In fact, this mechanism creates enormous order activities in markets as market participants try to "ping" for hidden liquidity inside of the spread by posting corresponding "fleeting orders" which are canceled a few instants later if they do not get executed.

This chapter aims at shedding light on the use of undisclosed orders in an opaque limit order market. To our best knowledge, this study is the first one providing empirical evidence on (entirely) hidden order placements in continuous trading of one of the largest equity markets worldwide. Specifically, using data from the NASDAQ TotalView message stream allows us to retrieve information on hidden depth. Employing an ordered response approach with censoring mechanism yields conditional probabilities of hidden order *locations* given the state of the market and provides insights on the distribution of hidden orders across different price levels. Hence, we perform statistical inference on the aggressiveness of hidden order placements (e.g., within the spread) and analyze how hidden liquidity submitters balance their risk of non-execution and compete for liquidity supply. Our findings based on a wide cross-section of NASDAQ stocks show that "dark" liquidity supply is significantly driven by market conditions and thus predictable in terms of the state of the (displayed) order book. We show that under certain market conditions, there is significant competition for hidden liquidity supply inducing a clear narrowing of the "hidden" (and thus effective) bid-ask spread. Conversely, in situations where the risk of being picked off becomes high, we observe a significant reduction in hidden order submitters' aggressiveness. Moreover, we provide novel insights into competition for hidden liquidity provision and hidden order placements in the presence of high-frequency trading.

The current tendency of trading platforms toward more opaqueness is observable on all major markets. We can differentiate between three major categories. The first group of markets, including various non-U.S. markets, such as the London Stock Exchange, Frankfurt Stock Exchange (XETRA), Australian Stock Exchange (ASX), Euronext, the Madrid Stock Exchange and the Toronto Stock Exchange, among others, offer the possibility of posting *only* iceberg orders (or so-called reserve orders) where the trader is obliged to show only a small proportion (called "peak") of the posted order size. The second category of trading platforms allows to use *both* reserve and hidden orders and thus offers the option to entirely hide an order. Prominent examples are NASDAQ, the New York Stock Exchange (NYSE), BATS (Best Alternative Trading System) – currently the third largest equity market in the U.S. – and the largest U.S. Electronic Communication Network (ECN) Direct Edge. According to the report by the Securities and Exchange Commission (2010), these markets cover approximately 75% of share volume in National Market System (NMS) stocks. The third group of modern trading systems are so-called dark pools where liquidity supply is hidden and no information on order matching and trading actions is provided to other market participants.

Recent empirical evidence shows that "dark trading" is not negligible and is increasingly popular. For instance, Bessembinder, Panayides, and Venkataraman (2009) report that 44% of order volume is hidden and 18% of incoming orders are reserve orders on Euronext Paris. Frey and Sandås (2009) show that reserve orders represent 9% of non-marketable orders with sizes of $12 - 20$ times the average in German XETRA trading. The Securities and Exchange Commission (2010) reports that 32 dark pools

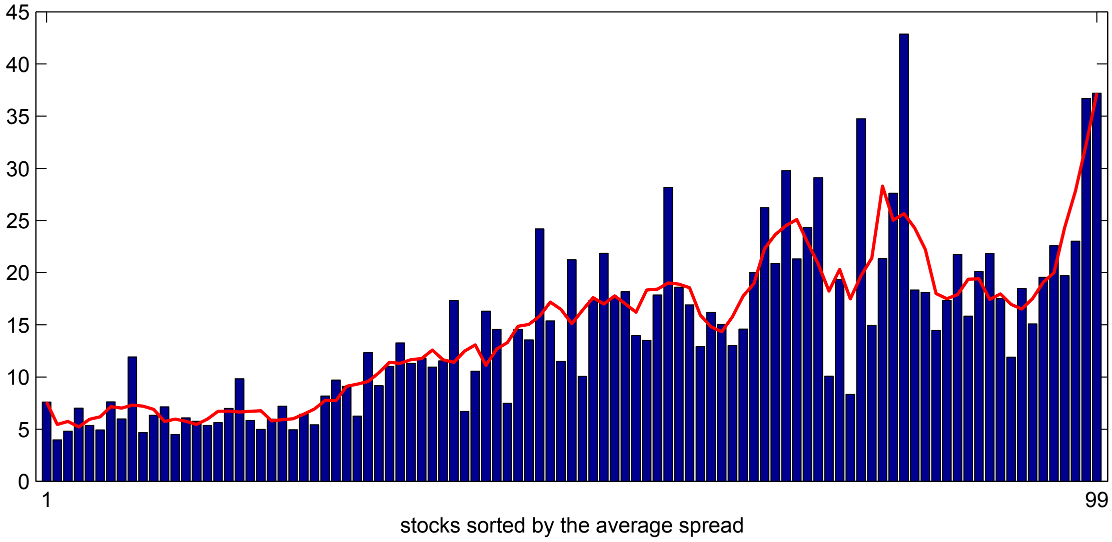


Figure 3.1: Percentage of trading volumes executed against hidden depth for 99 NASDAQ stocks representing a wide cross-section of the market. The stocks are sorted according to their average bid-ask spreads during the investigation period.

in the U.S. contribute approximately 8% of trading volume in NMS stocks. Figure 3.1 shows percentages of trading volume executed against hidden liquidity for 99 NASDAQ stocks used in our empirical analysis. Averaged across a wide range of the market, approximately 14% of the share volume originates from hidden depth. However, for some stocks, especially those revealing high spreads, it can be even greater than 40%.

The major motivation for hiding orders is to camouflage trading intention. The latter increases execution risk as the display of (large) orders may cause impatient traders to retreat (Moinas, 2010) and may lead to higher liquidity competition (Buti and Rindi, 2011). Moreover, posting limit orders induces front-running strategies (Harris, 1997), and the risk of adverse selection ("picking off risk"; see Harris, 1996). By hiding an order, execution risks can be reduced, while, on the other hand, the risk of non-execution rises as trading counter parties are not obviously attracted. Moreover, typically, hidden orders lose time priority to displayed orders. Hence, for a hidden order submitter it is crucial to balance the risk of non-execution vs. the risk of adverse selection.

Our empirical methodology is designed to provide insights into these motivations. To identify the locations of hidden orders, we employ two approaches. Firstly, we measure hidden order aggressiveness in terms of the distance between the order price and the best (visible) quote on the own side of the limit order book (LOB). The larger this distance the deeper a hidden order is placed within the spread and the higher is the order aggressiveness. The second approach employs the distance to the best visible quote on the opposite side of the market. The lower this distance, the lower the transaction costs for a market order submitter on the opposite side. We show that both distance measures are necessary to fully capture the determinants of hidden order placements. Using this setup, we analyze whether hidden order placements can be predicted using the observable state of the limit order book and can be explained by hidden order suppliers' motivation to balance execution risk and adverse selection risk. In particular, we address the major research questions: (i) Does hidden supply compete with observable order flow? (ii) Is there competition between hidden liquidity? (iii) How does hidden supply react to aggressive "pinging"?

Analyzing hidden order placements for 99 stocks covering a wide cross-section of the NASDAQ market in 2010, we can summarize the following results: First, hidden order placements follow trade directions in order to increase execution probabilities and to reduce adverse selection. In particular, market participants submit hidden orders less aggressively when the price moves in their favorable direction. Second, the "hidden" spread is positively correlated with the observed spread. This is particularly true for stocks with comparably high (average) spreads. Third, there is significant competition for the provision of liquidity. This is true for hidden liquidity as traders use more aggressive hidden orders after observing competing hidden depth on the own side. Moreover, it is also true for the competition between hidden and disclosed liquidity. The latter is empirically supported by a strong (positive) correlation between undisclosed orders and the visible depth on the same side of the market. Fourth, hidden order submitters become more defensive when high-frequency traders actively "ping" for undisclosed volume in the spread. Overall, our findings clearly show that hidden orders are placed strategically to balance non-execution and adverse selection risks.

The remainder of this chapter is organized in the following way: In Section 3.2, we review the theoretical and empirical studies related to undisclosed orders, and formulate our economic hypotheses. Section 3.3 briefly introduces the market environment, but discusses in details on the data construction and descriptive statistics. In Section 3.4, we introduce our econometric model for undisclosed orders. Section 3.5 documents our empirical findings. Section 3.6 concludes.

3.2 Economic Reasoning of Optimal Order Display

3.2.1 Theories on Undisclosed Orders

A major motivation for posting a limit order is to minimize transaction costs by appropriately choosing the limit price and to signal trading intention to other market participants in order to attract counterparties which might be not in the market yet (according to Harris (1996), so-called "passive traders"). Attracting a counterparty is important to increase the execution probability and to decrease the execution time of the position. However, signaling trading intention may induce adverse price effects. For instance, according to Moinas (2010), displaying large orders may cause "defensive" market order traders to retreat from the market as soon as they interpret the signal as inside information. Moreover, "parasitic" traders (Harris, 1997) may exploit the information value of a big order by using front-running strategies. Finally, posting a limit order induces the risk of being picked off and thus adverse selection (Harris, 1996) if price changes are stronger than expected and the order is sold (bought) too cheap (expensive).

Based on these economic reasoning, several theories on the usage of undisclosed orders have been developed. Esser and Mönch (2007) propose a static framework in which the trader optimizes the peak size and limit price of reserve orders by continuously monitoring and balancing exposure risk against execution risk. Moinas (2010) presents a theoretical model where informed traders, as well as large liquidity traders, use reserve orders to mitigate the information leakage. Cebiroglu and Horst (2011) propose a model where traders decide on the peak size of the iceberg order by accounting for the exposure-induced market impact. Buti and Rindi (2011) present a dynamic framework where the trader chooses her optimal strategy by simultaneously deciding on trading direction, aggressiveness, size and peak proportion of the order. To our best knowledge, it is the only theoretical model that explicitly incorporates the possibility of hidden order placements into traders' trading options.

3.2.2 Empirical Evidences on Undisclosed Orders

The empirical literature on reserve orders has been growing remarkably during the last decade, partially due to its proliferation in limit order markets and the increasing availability of data. Studying trading on Euronext Paris, Bessembinder, Panayides, and Venkataraman (2009) document that reserve orders induce lower implementation short fall costs but longer times to fill. De Winne and D'Hondt (2007) examine similar data and find that the detection of hidden depth increases order aggressiveness on the opposite side. Fleming and Mizrach (2009) examine data from BrockerTec, the leading interdealer ECN for U.S. Treasuries and documenting that the use of reserve orders varies considerably with the quantity of hidden depth increasing with price volatility. All studies show that the decision on using reserve orders is strongly related to prevailing market conditions, as characterized by the bid-ask spread, order book depth and prevailing volatility.

Studying data from the Australian Stock Exchange (ASX), Aitken, Berkman, and Mak (2001) find that reserve orders do not have a different price impact than visible limit orders. According to their results, the use of reserve orders increases with volatility and the average order value, while it decreases in tick size and trading activity. Frey and Sandås (2009) analyze the Deutsche Börse's trading platform XETRA and show that the price impact of reserve orders depends on the executed fraction of its size with profitability increasing in the hidden proportion. Based on data from the Spanish Stock Exchange Pardo Tornero and Pascual (2007) find no significant price impact associated with the execution of hidden parts of reserve orders. These findings support the hypothesis that liquidity suppliers use reserve orders to compete for liquidity provision while preventing picking-off risks.

Tuttle (2006) shows that overall market depth increased significantly after NASDAQ introduced undisclosed orders. Moreover, they provide evidence for hidden sizes being predictive for future market price movements while the visible size conveys only little information. Anand and Weaver (2004) examine the abolition in 1996 and re-introduction in 2002 of reserve orders on the Toronto Stock Exchange and show that the spread and visible depth remain widely unchanged after both events. However, total depth, including both visible and hidden volume, significantly increases after the re-introduction. Both studies show that market quality is improved after the introduction of reserve orders and that informed traders tend to use them primarily to reduce the price impact.

3.2.3 Testable Hypotheses

Market microstructure theory (see e.g., Easley, Kiefer, and O'Hara, 1997) suggests that large bid-ask spreads reflect a high uncertainty of liquidity suppliers on the state of the market and the fundamental value of the asset. If hidden liquidity suppliers are dominantly liquidity-motivated, they will be reluctant to submit aggressive hidden orders inside of the spread as a high uncertainty also implies a high picking-off risk. Consequently, hidden depth inside of the spread will decrease if spreads widen. Conversely, for informed market participants who want to camouflage their informational advantage, wider spreads open more room for aggressive hidden order placements. Hence, in such a situation we would expect opposite effects with rising hidden liquidity placed inside the spread. Hence:

Hypothesis 1.A The probability of hidden depth inside the spread decreases with the size of the spread. (Liquidity traders use hidden orders)

Hypothesis 1.B The probability of hidden depth inside the spread increases with the size of the spread. (Informed traders use hidden orders)

Traders can use undisclosed orders to compete for the provision of liquidity while preventing others from undercutting their orders. Buti and Rindi (2011) demonstrate that undisclosed orders are part of equilibrium strategies of liquidity suppliers. In particular, when the depth on the own side of the market is high, traders prefer to place more aggressive hidden orders inside of the spread to increase the execution probability. Likewise, in case of a high depth on the opposite side of the market, hidden order submitters may place their orders deeply inside the spread in order to "win" the (own-side) competition for liquidity provision and to maximize the execution probability if an impatient trader from the opposite side posts a market order and crosses the spread. However, this strategy implies a high adverse selection risk as a high depth on the opposite side may reflect significant price pressure. Consequently, aggressive hidden orders placed deeply inside the spread are likely to be picked up if the price pressure piling up on the opposite side materializes. Accordingly, we might expect a(n) reduction (increase) of hidden order aggressiveness if hidden order submitters' exposure to adverse selection risk (risk of non-execution) dominates. Expecting a higher importance of non-execution risk, we formulate the following hypotheses:

Hypothesis 2.A The probability of hidden bid (ask) depth inside spreads increases when the bid (ask) visible depth is thick.

Hypothesis 2.B The probability of hidden bid (ask) depth inside spreads increases when the ask (bid) visible depth is thick.

Traders' order submission strategies depend not only on the current state of the limit order book but also on recent price movements and trading signals. The dynamic equilibrium model on visible order flow proposed by Parlour (1998) shows a "crowding out" effect among market orders: the probability of incoming sell (buy) market orders is lower after observing a buy (sell) market order which is in line with the well-known strong persistence in trade directions. This effect implies that visible bid (ask) limit orders have a higher execution probability after a sell (buy) market order. This hypothesis is supported by Hall and Hautsch (2005) showing that price movements are positively (negatively) correlated with the aggressiveness of visible buy (sell) limit orders. We expect that liquidity suppliers take advantage of these trading signals by posting hidden orders deeper inside the spread and thus increase execution probabilities. However, as argued above, in situations where liquidity suppliers aim at benefiting from price pressure built up on the opposite side of the market, their exposure to adverse selection risk increases, which – in the extreme case – might dominate. A similar scenario arises in situations of prevailing price changes. Also here, it might be advantageous to place an aggressive hidden ask (bid) order after observing a price increase (decrease) and thus to serve as a counter party if further buy (sell) market orders are expected. Again, a counter-effect arises by an increased picking-off risk. Assuming that traders tend to minimize picking-off risk rather than non-execution risk, we expect hidden bid (ask) depth to decrease (increase) in case of sell (buy) pressure.

Moreover, recent trading and price signals may not only generate activity on the opposite side of the market but may also trigger liquidity competition on the own side. If liquidity suppliers expect momentum in prevailing price movements and, according to the crowding out hypothesis, persistence in trade directions, (hidden) liquidity competition on the same side should become stronger because liquidity suppliers face a higher risk of non-execution. Hence, for order activities on the own side, we expect protection against non-execution risk to dominate leading to two testable hypotheses:

Hypothesis 3.A The probability of hidden bid depth inside of the spread decreases (increases) when the prevailing trade is seller (buyer)-initiated. The converse effect applies for hidden ask depth.

Hypothesis 3.B The aggressiveness of hidden bid depth increases (decreases) when the price moves up (down).

Traders' decision on using undisclosed orders might also depend on the asset's volatility. Foucault (1999) shows that volatility is an important parameter in order submission strategies. Indeed, higher volatility implies higher uncertainty on the value of the asset and thus increases the picking off risk. This mechanism is true for visible orders but should similarly also apply to hidden orders:

Hypothesis 4 The aggressiveness of hidden depth is negatively correlated with prevailing asset price volatility.

Hidden depth is a priori unobservable but is ex post identifiable as soon as it gets executed. This is most clearly seen if a limit order posted inside of the spread gets immediate execution. Such information gives hidden liquidity providers hints on the prevailing competition for hidden liquidity supply which in turn affects their submission strategies. As a result of higher (hidden) liquidity competition, they would post more aggressive orders to increase their execution probability. Moreover, Buti and Rindi (2011) show that detections of hidden depth encourage more undisclosed order submissions as long as picking-off risks (and thus adverse selection risks) do not become too high. The reasoning is that market participants interpret the detection of hidden volume as a signal of high liquidity demand and compete for supplying it. Accordingly, we postulate the following hypothesis:

Hypothesis 5 The aggressiveness of hidden bid (ask) depth increases after some hidden bid (ask) depth has been executed.

In modern trading, high-frequency trading (HFT) plays an increasingly important role (see e.g., Angel, Harris, and Spatt, 2010; Securities and Exchange Commission, 2010) and might also influence the supply for hidden liquidity. In fact, HFT algorithms typically employ directional strategies, such as quote-matching and fast-trading ("scalping"), i.e., the idea to post a limit order in front of some other limit order which is expected to reveal information and to pick up the mis-priced limit orders quickly when information flows in. To avoid such effects, traders prefer using undisclosed orders. However, some HFT trading algorithms also embed strategies for detecting hidden depth, such as "pinging", where visible limit orders are posted in the spread in order to test whether they might get executed. Our empirical results show that such effects create enormous order activities on NASDAQ. Pinging strategies, combined with scalping, induce severe picking-off risks for undisclosed orders and may make them quite inefficient. Indeed, Buti and Rindi (2011) theoretically show that when hidden depth can be perfectly detected there is no ground for traders using undisclosed orders to reduce exposure risks. Accordingly, we expect that hidden liquidity suppliers become less aggressive if high-frequency traders become very active in the market:

Hypothesis 6 The aggressiveness of hidden depth decreases as HFT activities on the opposite side of the market increase.

In Moinas's (2010) theoretical framework, informed traders use undisclosed orders to mitigate information leakage. Typically, information asymmetry is highest during the opening period as overnight information has to be processed. Accordingly, we expect a higher hidden order aggressiveness in this period compared to the rest of the trading day. Moreover, Esser and Mönch (2007) show that traders tend to display more of order sizes when they approach trading closure. This is driven by the typical desire to close a position before the end of the trading session. Accordingly, trading intentions are revealed such that order execution probabilities are increased due to a higher time priority of visible orders. Buti and Rindi (2011) also argue that reserve orders are preferable to hidden orders in their framework when the time horizon becomes shorter. This leads to the following hypotheses:

Hypothesis 7.A The aggressiveness of hidden depth is higher during the opening hour.

Hypothesis 7.B The aggressiveness of hidden depth is lower during the closure hour.

3.3 Measuring Hidden Order Locations

3.3.1 Institutional Background

As one of the largest electronic limit order markets in the world, the NASDAQ Single-Book platform provides an unified procedure for passing limit orders from ECNs (Brut and INET) and the traditional dealer-quote system. In particular, it treats a market maker's quote as a pair of limit orders on both sides of the market and aggregates them into a centralized order book. During continuous trading between 9:30 and 16:00 Eastern Time, the system matches incoming orders against the best (in term of price) prevailing (possibly undisclosed) orders in the LOB. If there is insufficient volume to fully execute the incoming order, the remaining part will be consolidated into the LOB. Besides limit orders and market orders, NASDAQ provides both reserve orders and hidden orders.[1] As a reward for traders disclosing their orders, the hidden part of undisclosed orders loses time priority compared to visible limit orders or peaks of reserve orders on the same price level. Market makers at NASDAQ may also provide hidden depth. The NASDAQ Stock Market trading rule (NASDAQ, 2008) requires the market maker to display at least one round lot size. In this case, the market maker's quotation corresponds to a pair of reserve orders.

3.3.2 Data

We conduct our study based on 99 stocks traded on NASDAQ over the period of October 2010 corresponding to 21 trading days. To represent a wide cross-section across the market, we select stocks according to market capitalization. We first rank the 500 biggest stocks according to their market capitalizations recorded in the Center for Research in Security Prices (CRSP) database on 30th September 2010. Then, we restrict the sample by selecting a stock out of every percentile resulting in 99 stocks which are divided into three equal-size groups according to their average spreads and trade frequencies.

We retrieve historical NASDAQ market conditions from TotalView-ITCH data. NASDAQ TotalView^{SM} data, surpassing NASDAQ Level 2, is the current standard NASDAQ data feed for displaying the real-time full order book depth for market participants. Historical data files record rich information on order activities, including limit

[1]NASDAQ also provides so-called "discretionary orders" with a displayed price and size as well as a non-displayed discretionary price range. When the discretionary price range is hit by a matching order, the discretionary order converts into an IOC (Immediate or Cancel) market order. This order type also allows to hide trading intention. However, we do not consider discretionary orders as undisclosed orders because (i) they *take* liquidity rather than providing it, and (ii) it is very difficult to identify them using TotalView-ITCH data as HFT algorithms generate an enormous number of IOC orders.

Table 3.1

Cross-Sectional Summary Statistics on the Characteristics of the Selected Stocks.

The sample consists of 99 NASDAQ stocks during the period of October 2010 with 21 trading days. We divide them into three equal-size groups according to the average spread (AvgSpr) and the number of trades (AvgTrd). For each group, we report summary statistics of the following variables: MktCap is the market capitalization recorded in CRSP at the end of trading on 30th September, 2010. AvgSpr (in ¢) is the average spread in dollar cent. AvgTrd is the average number of daily trades. AvgHit is the average number of daily trades (partially or totally) traded against hidden depth. AvgHit (in %) is the average percentage of daily trades (partially or totally) traded against hidden volume. AvgVol is the average daily trading volume (in thousand shares). AvgHVol is the average daily trading volume traded against hidden depth. AvgHVol (in %) is the average daily percentage of executed hidden volume relative to overall trading volume.

		MktCap (in bil. $)	AvgSpr (in ¢)	AvgTrd	AvgHit	AvgHit (in %)	AvgVol ($\times 10^3$Shr)	AvgHVol ($\times 10^3$Shr)	AvgHVol (in %)
Entire	Mean	7.83	5.38	1861	429	20.1	3.93	0.59	14.6
Sample	Median	2.16	3.77	1083	178	18.7	2.05	0.21	13.5
	Std.	28.19	6.01	2616	959	7.9	5.87	1.59	8.1
	Min.	0.89	1.07	98	12	9.1	0.12	0.01	3.9
	Max.	259.90	34.91	20583	8446	46.0	41.37	14.37	42.8
AvgSpr	Small	9.00	1.36	2199	316	14.0	5.84	0.37	6.9
Groups	Medium	3.81	3.68	1776	414	20.0	3.08	0.54	15.0
(means)	Large	10.69	11.10	1608	558	26.2	2.88	0.86	21.8
AvgTrd	Low	1.50	8.85	461	91	20.4	0.70	0.10	16.3
Group	Medium	2.91	4.05	1121	202	18.2	2.10	0.24	12.7
(means)	High	19.09	3.23	4002	994	21.6	9.00	1.42	14.7

order submissions, cancellations, executions of visible and hidden orders as well as a unique identification number for every (visible) limit order and peak of reserve orders.

We reconstruct the historical LOB using the algorithm proposed by Huang and Polak (2011). Their algorithm continuously updates the LOB according to all reported messages and represents the exact state of the LOB as shown to TotalView subscribers in real time. Furthermore, we identify the attribute of a limit order (cancelled or filled) and compute its lifetime by tracking it through its order ID.[2] Finally, we aggregate sequences of executions of buy (sell) limit or hidden orders occurring in less than 0.5 seconds into one sell (buy) market order. If a limit order is recorded immediately after such a sequence, it is also aggregated with the entire sequence being considered as a marketable limit order. Finally, to avoid erratic effects during the market opening and closure, our sample period covers only the periods between 9:45 and 15:45.

Table 3.1 summarizes major characteristics of the selected stocks. They cover a wide universe of stocks with market capitalization ranging from 900 million to 260 billion US dollar. We find clear evidences for a high popularity of undisclosed orders in NASDAQ

[2]Limit order book reconstruction and limit order tracking is performed by the software "LOBSTER" which can be accessed at `http://lobster.wiwi.hu-berlin.de`.

Table 3.2

Cross-Sectional Summary Statistics on Limit Order Executions and Cancellations

The sample consists of 99 NASDAQ stocks during the period of October 2010, with 21 trading days. We divide them into three equal-size groups according to the average spread (AvgSpr) and the number of trades (AvgTrd). For each group, we report cross-sectional summary statistics for the following variables: NumLO is the average daily number of limit orders (including peaks of reserve orders). NumCanc is the average daily number of limit order cancellations before getting (partially) executed. MedCTim is the median of the lifetime of canceled visible limit orders. MedETim is the median of the lifetime of executed limit orders. VWETim is the volume-weighted execution time of limit orders. NumALO is the average daily number of limit orders placed inside the spread (aggressive limit orders). NumACan (in %) is the average daily percentage of cancelled aggressive limit orders placed inside the spread. AvgATim is the average lifetime of cancelled aggressive limit orders.

		NumLO ($\times 10^3$)	NumCanc (in %)	MedCTim (sec.)	MedETim (sec.)	VWETim (sec.)	NumALO ($\times 10^3$)	NumACan (in %)	AvgATim (sec.)
Entire	Mean	57.92	94.7	9.7	12.9	142.0	3.84	76.3	3.11
Sample	Median	29.41	94.8	9.2	10.7	103.7	2.52	79.7	2.12
	Std.	84.50	1.9	6.3	9.5	152.2	5.81	14.9	4.79
	Min.	5.01	90.9	0.0	0.8	39.9	0.07	29.4	0.02
	Max.	650.66	99.2	33.2	60.3	981.2	49.9	98.7	41.0
AvgSpr	Small	79.93	93.7	10.0	15.8	154.8	1.85	61.1	1.27
Groups	Medium	47.14	94.5	11.2	10.6	114.2	4.06	79.6	3.23
(means)	Large	46.69	95.9	7.8	12.4	156.9	5.62	88.1	4.84
AvgTrd	Low	14.67	96.0	12.5	19.9	216.2	2.42	85.1	2.1
Group	Medium	33.50	94.3	9.3	12.2	109.5	2.50	75.5	2.04
(means)	High	125.6	93.9	7.2	6.7	100.1	6.61	68.2	5.12

trading. On average, approximately 15% of the trading volume and 20% of all trades are executed against hidden depth. The average size of executed hidden depth is slightly smaller than that of visible depth. This is partially due to active HFTs who use high-speed hidden depth detecting algorithms to compete for trading against hidden volume. Moreover, note that only a small proportion of existing hidden depth gets executed (see e.g., Bessembinder, Panayides, and Venkataraman, 2009; Frey and Sandås, 2009). Hence, the share of (undetected) hidden depth is much greater than the magnitudes reported in the table. Furthermore, we show that the proportion of trading volume executed against hidden depth increases as the (average) spread becomes wider. Hence, trades of high-spread stocks are more likely to get price improvements.

Table 3.2 reports summary statistics on limit order executions and cancellations. We find that on average approximately 95% of all limit orders are cancelled without getting (partially) executed. This strikingly high number is robust across the sample with the cross-sectional standard deviation being very low. In fact, the stock with the smallest proportion of cancellations still reveals a percentage of 91%. Conversely, we observe the most extreme situation of a stock revealing 99% cancellations. Moreover, the median lifetime of cancelled orders is less than 10 seconds. For limit orders placed inside the spread, the average time until cancellation is even just around 3 seconds. This effect is obviously driven by a strong influence of HFT-induced pinging strategies aiming at detecting hidden orders inside the spread. Interestingly, large visible limit orders have much longer execution times than small orders. This is indicated by the volume-weighted execution time of 142 seconds being substantially higher than the median lifetime of executed limit orders (12.9 seconds). This evidence is in line with extant empirical studies of the market impact of limit orders (see e.g., Eisler, Bouchaud, and Kockelkoren, 2011; Hautsch and Huang, 2012b), showing supportive evidence of large traders' economic motivation for using undisclosed orders. Finally, cancellation rates of aggressive limit orders turn out to be lower as they have higher execution probabilities.

3.3.3 Identifying Undisclosed Orders

It is in the nature of things, that information on hidden order placements is not provided by an exchange. Therefore, from classical transaction data sets, as, e.g., the Trade and Quote (TAQ) database released by the New York Stock Exchange (NYSE), it is impossible to infer on hidden orders. This difficulty is the major reason for the lacking empirical evidence on hidden order placements. Message data, as provided by TotalView, however, contain information on *any* activities affecting the *visible* part of the LOB. In particular, it specifically reports *executions* against hidden orders which allow to identify the exact position of hidden depth in the LOB. As illustrated below, these details can indeed be utilized to conduct statistical inference on undisclosed order submissions.

In general, we can distinguish between trading scenarios where we can *distinctly* (ex post) identify the location of hidden volume and situations where we can isolate at least *partial* information on the existence of undisclosed volume. Figure 3.2 illustrates an example of the first scenario where the best (visible) quotes in the LOB are 24.86 (bid) and 24.91 (ask) before a buy limit order with limit price 24.91 is posted. As there is a hidden ask order at price 24.90 inside of the spread, the incoming order is firstly partially filled by this order resulting in a type "P" trade message (denoting executions against hidden depth in the NASDAQ ITCH 4.0 format). Next, the remaining part of the buy order is executed against the visible depth at the best ask resulting in an "E" message. Finally, hidden depth at the best (visible) ask gets executed resulting in a further "P" message. The remaining (non-executed) part of the incoming order enters the book as a new buy limit order submission (type "A" message) at 24.91.[3]

This example shows that due to the existence of hidden depth, the market order submitter faces a better execution price than expected from the visible LOB. If the trader is able to predict the existence of hidden depth within the spread, she can incorporate these transaction cost savings in her trading strategy. Moreover, it is illustrated that the visible depth has execution priority over the hidden depth at the same price, no matter *when* the order has been placed. Hence, if further depth on the best ask level cumulates,

[3] As in general, visible and hidden volumes are provided by more than one order at the same price, we typically observe a sequence of simultaneous "P" and "E" messages.

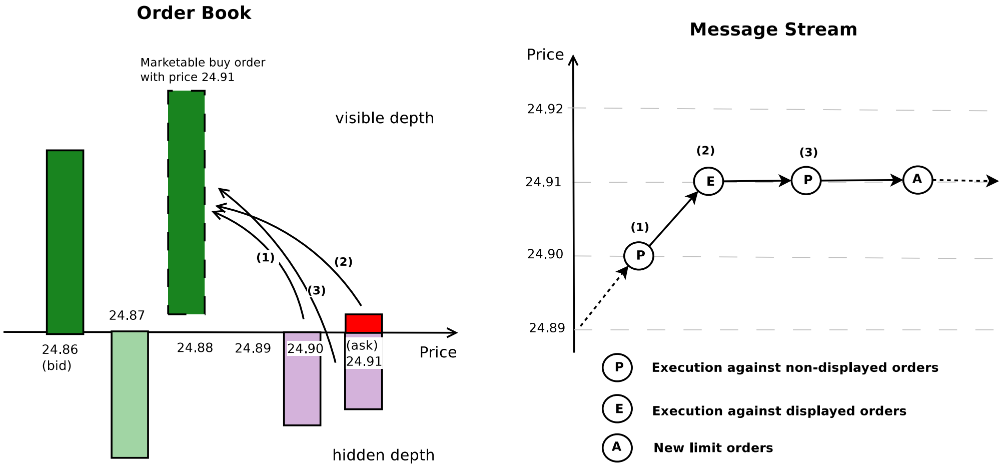


Figure 3.2: Left: Stylized trading scenario in a limit order book where a buy market order is executed against hidden volume on the ask side and is uniquely identified. Bid orders are marked by green, whereas ask orders are marked by red. All orders above the horizontal axis are visible, whereas orders below the axis are hidden. The numbered arrows indicate the matching process. Right: Sequence of generated messages (in NASDAQ ITCH 4.0 format) resulting from this transaction.

the time-to-fill of any hidden order becomes longer. Finally, since the execution of the hidden part is uniquely identified, we can exactly locate the undisclosed order in this example.

Figure 3.3 shows a scenario which allows to extract at least incomplete information on hidden order placements. Suppose a buy limit order is submitted inside of the spread with price 24.88. The fact that the limit order does not get executed (otherwise we would have been observed a "P" message), reflects that there cannot be any hidden ask volume posted on a price level lower than 24.89. Hence, this observation reveals information about *non-existence* of hidden depth. We refer to such an observation as *censored* as it only provides a lower (upper) bound for the location of hidden ask (bid) volume.

Finally, as illustrated by Figure 3.4, there might be a scenario where a marketable order is executed against two (or several) levels of visible depth. The fact that not even a part of the order is executed against hidden volume indicates the non-existence of hidden ask depth on any level up to (including) price level 24.90. Hence, also this observation is *censored* in the sense that it only yields a location bound.

Summarizing, we infer price information on undisclosed orders based on the following three scenarios:

1. Submission of a marketable order when the spread is larger than 1 tick. If the order gets executed at a price better than the corresponding best (visible) quote, we can exactly identify the hidden order location and thus obtain an "uncensored" observation.

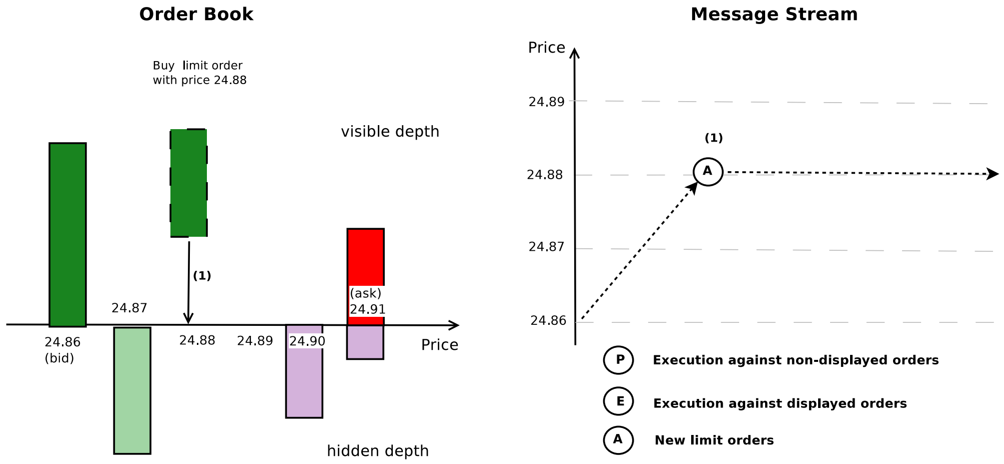


Figure 3.3: Left: Stylized trading scenario in a limit order book where a limit order placed into the spread reveals (partial) information about the hidden depth. Bid orders are marked by green, whereas ask orders are marked by red. All orders above the horizontal axis are visible, whereas orders below the axis are hidden. Right: Sequence of generated messages (in NASDAQ ITCH 4.0 format) resulting from this submission.

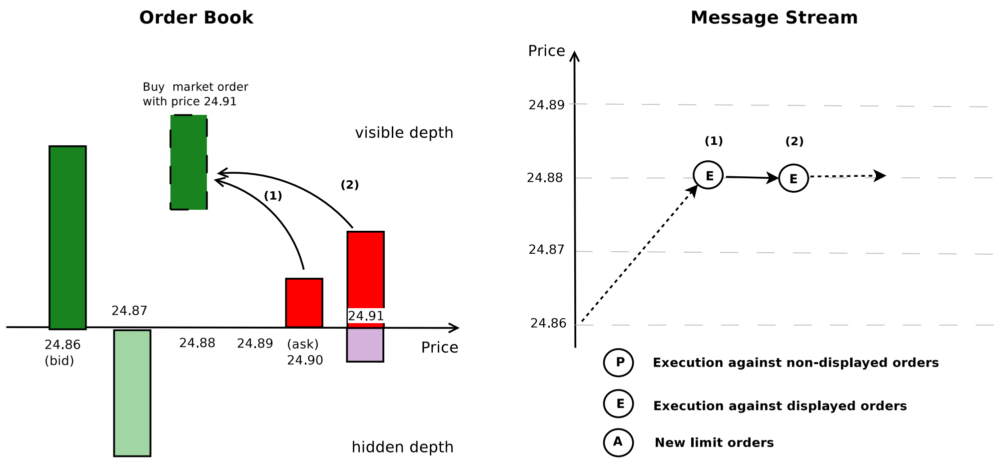


Figure 3.4: Left: Stylized trading scenario in a limit order book where a buy market order is executed against visible volume only and thus reveals (partial) information about the hidden depth. Bid orders are marked by green, whereas ask orders are marked by red. All orders above the horizontal axis are visible, whereas orders below the axis are hidden. Right: Sequence of generated messages (in NASDAQ ITCH 4.0 format) resulting from this submission.

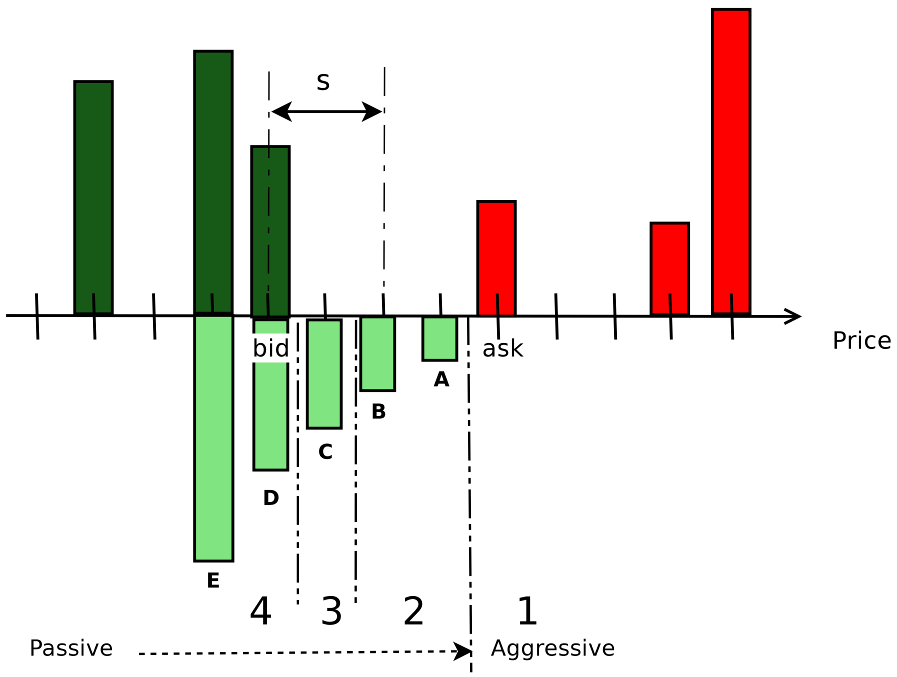


Figure 3.5: Graphical illustration of the hidden order aggressiveness measure s and corresponding classifications for the case of large-spread stocks (4 categories).

2. Submission of a limit order inside of the spread. If it is not executed, we certainly know that there is undisclosed order with better limit price. This results into a "censored" observation.

3. Submission of a marketable order with size greater than the depth at the corresponding best (visible) quote. As this order may be split across several levels, we can infer on hidden depth at- or behind-the-market. The observation can be uncensored or censored depending on whether it is partially filled by hidden depth or not.

3.3.4 Measuring the Aggressiveness of Undisclosed Orders

Biais, Hillion, and Spatt (1995) classify the aggressiveness of a limit order by measuring its (price) distance to the prevailing best quotes. This scheme has been widely employed in the empirical literature on limit orders (e.g. Griffiths, Smith, Turnbull, and White, 2000) and reserve orders (e.g., Bessembinder, Panayides, and Venkataraman, 2009). Accordingly, we measure the aggressiveness of undisclosed orders in a similar way. In particular, we measure distances of hidden order placements relative to best quotes on the own and opposite side of the market.

Let p^a and p^b denote the best ask and bid quote and p^o represents the limit price of the undisclosed order. According to our first scheme, we measure the distance of the undisclosed order to the best quote on the order's own side,

$$s = \begin{cases} p^o - p^b & \text{for undisclosed buy orders,} \\ p^a - p^o & \text{for undisclosed sell orders.} \end{cases}$$

Hence, the larger s, the deeper the order is placed in the spread. Conversely, if $s \leq 0$, the undisclosed order is placed in the book and can be either a reserve order or a hidden order. Accordingly, s measures aggressiveness from the liquidity supplier's perspective. Due to the fact that most observations only reveal incomplete, i.e. "censored", information, it is most natural to measure hidden order aggressiveness in terms of categories. As discussed in the following sections, this allows for straightforward and computationally tractable econometric modelling avoiding severe assumptions on the functional form. Depending on the underlying (average) size of the spread, we choose different categorization schemes. In particular, we divide the set of hidden order locations into 2, 3 and 4 categories for small-spread, medium-spread and large-spread stocks, respectively. Table 3.3 gives the chosen categories depending on s. The choice of the groups is motivated by the need to have a sufficient number of observations in each category and to use a preferably fine categorization *within* the spread. Figure 3.5 illustrates the resulting scheme for the case of large-spread stocks.

As bid-ask spreads are not constant over time, the distance measure s is not sufficient to fully capture hidden order locations. It is rather necessary to measure orders' aggressiveness also in terms of the distance to the opposite side of the market. Accordingly,

$$d = \begin{cases} p^a - p^o & \text{for undisclosed buy orders,} \\ p^o - p^b & \text{for undisclosed sell orders.} \end{cases}$$

Hence, d represents undisclosed orders' aggressiveness from the liquidity demander's perspective, see Figure 3.6. The smaller d, the lower the actual transaction costs for a market order submitter who gets executed against this undisclosed order. Note that d cannot become negative as any placement *behind* the opposite side of the market would immediately result into an execution. As shown by Table 3.3, we categorize d in a similar way to s. However, as d highlights the implied transaction costs induced by execution against undisclosed orders, we choose a categorization which is particularly fine close to the opposite side.

Note that the categorizations underlying the two measures can be partially overlapping. For instance, category 2 in Figure 3.5 overlaps with the categories 1 and 2 in Figure 3.6, while category 3 in Figure 3.6 overlaps with categories 3 and 4 in Figure 3.5. As shown in the empirical part of this chapter, this overlapping structure is particularly advantageous as it enables us to capture manifold changes of the hidden depth distribution by means of relatively simple and robust models.

Table 3.3 summarizes information on undisclosed orders. Firstly, the number of order submissions revealing hidden depth information is huge, especially for large-spread stocks. For these stocks, we observe on average approximately $113,000$ submissions.

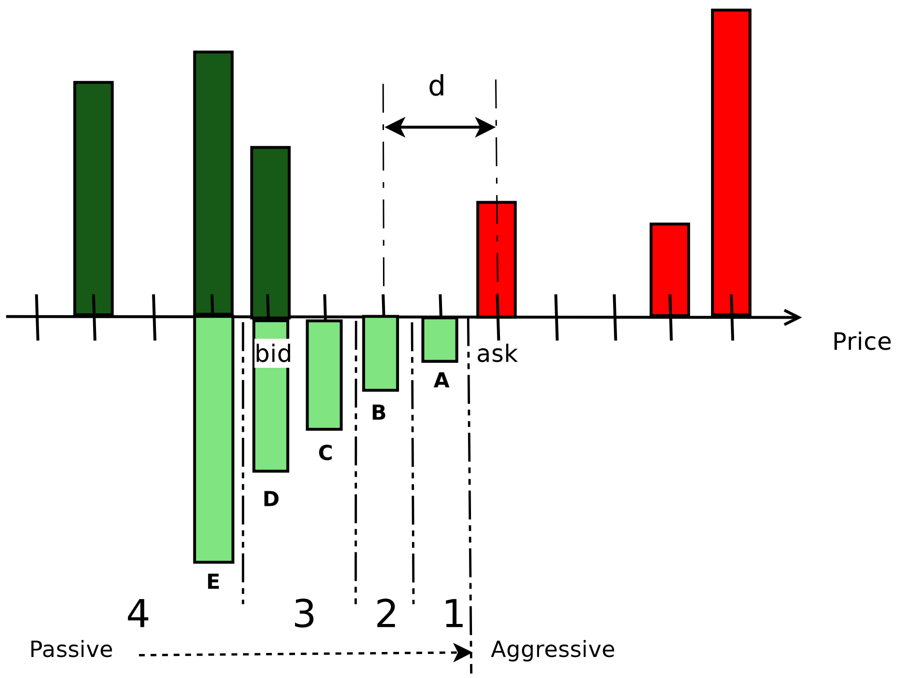


Figure 3.6: Graphical illustration of the hidden order aggressiveness measure d and corresponding classifications for the case of large-spread stocks (4 categories).

Table 3.3

Cross-sectional summary statistics on observations on undisclosed orders

The sample consists of 99 NASDAQ stocks during the period of October 2010 with 21 trading days. We divide them into three equal-size groups according to the average spread. The aggressiveness of undisclosed orders is measured by s and d as described in Section 3.3.4. We employ two, three and four categories for small-spread, medium-spread and large-spread stocks, respectively. Censored observations are defined as in Section 3.3.3. For each group we show cross-sectional statistics on total numbers (over all trading days).

Cate-	Distance	# Observation ($\times 10^3$)			Censored Obs. (%)			% Buy Orders (%)		
gory	(ticks)	max.	mean	min.	max.	mean	min.	max.	mean	min.
Aggressiveness measured by the distance to the own side quote (s)										
Spread group: small										
"1"	$s_t > 0$	147.65	29.32	1.71	96.5	83.3	26.2	52.7	49.8	45.5
"2"	$s_t \leq 0$	15.33	3.42	0.46	50.6	19.9	5.5	62.7	50.2	43.3
Spread group: medium										
"1"	$s_t > 1$	357.27	50.02	7.20	98.8	96.0	92.6	57.0	50.0	44.1
"2"	$s_t = 1$	91.33	21.22	3.91	97.0	87.1	65.1	56.9	49.9	44.2
"3"	$s_t \leq 0$	63.17	7.895	1.12	83.3	65.0	41.2	58.4	48.5	43.6
Spread group: large										
"1"	$s_t > 3$	623.82	76.58	12.13	99.8	97.8	90.6	56.1	50.5	44.7
"2"	$s_t = 2, 3$	193.84	17.04	0.40	98.6	86.4	38.2	56.8	51.2	47.7
"3"	$s_t = 1$	107.46	9.48	0.33	90.2	64.7	23.0	56.0	49.0	43.6
"4"	$s_t \leq 0$	121.38	10.03	0.77	94.9	81.3	60.0	57.3	49.3	42.1
Aggressiveness measured by the distance to the opposite side quote (d)										
Spread group: small										
"1"	$d_t = 1$	156.96	31.12	2.07	94.1	75.2	22.4	54.4	49.8	45.1
"2"	$d_t > 1$	15.19	1.62	0.10	98.0	87.5	69.6	71.5	51.3	37.2
Spread group: medium										
"1"	$d_t = 1$	327.35	57.75	9.89	97.5	93.2	86.3	55.6	50.0	45.0
"2"	$d_t = 2$	80.48	11.27	1.97	95.8	88.1	77.9	54.9	48.9	40.9
"3"	$d_t > 2$	99.99	10.11	0.94	93.7	81.3	52.5	61.0	50.4	39.2
Spread group: large										
"1"	$d_t = 1$	561.97	61.63	12.96	99.8	97.2	92.1	55.4	50.4	42.5
"2"	$d_t = 2$	174.53	17.44	2.89	99.7	93.9	82.7	55.8	50.0	39.6
"3"	$d_t = 3, 4$	162.94	16.73	1.51	99.3	89.6	68.4	56.3	49.9	41.1
"4"	$d_t > 4$	147.06	17.34	0.64	98.0	82.0	46.4	77.4	52.0	41.7

Table 3.4
Definitions of LOB control variables hidden orders on the buy side
"Aggressive limit orders" are defined as limit orders undercutting the prevailing best quote. "Fleeting orders" are defined as limit orders that are canceled within one second after the submission.

SPR	$\equiv$ log(best ask/best bid)
DPO	$\equiv$ log(depth at best bid)
DPG	$\equiv$ log(depth at best ask)
TYP	$\equiv$ 1 if the prevailing trade is seller-initiated; -1 otherwise
RET	$\equiv$ log return over the prevailing 5 minutes
VOL	$\equiv$ market price range (maximum - minimum) over the prevailing 5 minutes
HVO	$\equiv$ log(1+volume of executed hidden bid depth during the prevailing 1 minute)
HVO_5	$\equiv$ log(1+volume of executed hidden bid depth during the prevailing 5 minutes)
HRO	$\equiv HVO - HVO_5$
HVG	$\equiv$ log(1+volume of executed hidden ask depth during the prevailing 1 minute)
HVG_5	$\equiv$ log(1+volume of executed hidden ask depth during the prevailing 5 minutes)
HRG	$\equiv HVG - HVG_5$
ALO	$\equiv$ log(1+ number of aggressive buy limit orders that are not canceled during the prevailing 3 minutes)
ALG	$\equiv$ log(1+ number of aggressive sell limit orders that are not canceled during the prevailing 3 minutes)
HFO	$\equiv$ log(1+ number of fleeting buy orders during the prevailing 3 minutes)
HFG	$\equiv$ log(1+ number of fleeting sell orders during the prevailing 3 minutes)
OPN	$\equiv$ 1 trading before 10 : 30; 0, otherwise.
CLS	$\equiv$ 1 trading after 15 : 00; 0, otherwise.

Secondly, more than 90% of all observations are censored in the sense of reflecting only the upper bound of aggressiveness of hidden depth. Thirdly, the number of observations on the buy and sell side are very similar.

Finally, note that we label the underlying categories in a consistent way with the least aggressive categories being associated with the highest label and the most aggressive category being associated with the lowest label.

3.3.5 Capturing Market Conditions

To test our postulated hypotheses and to relate the usage of undisclosed orders to prevailing market conditions, we construct different variables representing various states of the market. Table 3.4 gives the exact definitions of constructed variables used for hidden order submissions on the buy side.

For statistical inference on the sell side, we modify some of the variables as follows:

DPO ≡ log(depth at best ask)
DPG ≡ log(depth at best bid)
TYP ≡ 1 if the prevailing trade is buyer-initiated; −1 otherwise
RET ≡ negative log return over the prevailing 5 minutes
HVO ≡ log(1+volume of executed hidden ask depth during the prevailing 1 minute)
HVO_5 ≡ log(1+volume of executed hidden ask depth during the prevailing 5 minutes)
HVG ≡ log(1+volume of executed hidden bid depth during the prevailing 1 minute)
HVG_5 ≡ log(1+volume of executed hidden bid depth during the prevailing 5 minutes)
ALO ≡ log(1+ number of aggressive sell limit orders that are not canceled during the prevailing 3 minutes)
ALG ≡ log(1+ number of aggressive buy limit orders that are not canceled during the prevailing 3 minutes)
HFO ≡ log(1+ number of fleeting sell orders during the prevailing 3 minutes)
HFG ≡ log(1+ number of fleeting buy orders during the prevailing 3 minutes)

The prevailing LOB state is represented by the visible bid-ask spread (SPR), reflecting the transaction costs of immediate trading, the visible depth on the best level on the own side (DPO) and the visible depth on the best level on the opposite side (DPG). To capture the impact of prevailing trade signals, we include a dummy variable (TYP) representing the most recent trading direction and the prevailing five-minute mid-quote return (RET) capturing short-term price movements. Moreover, local price volatility (VOL) is included in terms of the (max/min) range of trade prices during the last 5 minutes. Information on hidden depth is incorporated by the short-run executed hidden depth on the own side and the opposite side (HVO, HVG), representing how successfully traders have detected pending hidden depth. Moreover, to assess the relative intensity of temporary hidden order executions, we compute the executed hidden depth during the last minute relative to that executed during the last five minutes (HRO, HRG). Moreover, HFT activities are captured by two variables, HFO and HFG, which are the number of fleeting orders on the own side and the opposite side, respectively. Defined as in Hasbrouck and Saar (2009), a "fleeting order" is a limit order that is canceled within one second after the submission and thus is posted to "test" for the existence of hidden volume. Using the intensity of fleeting orders as a proxy for HFT activities is inspired by Hendershortt, Jones, and Menkveld (2010). To differentiate between fleeting orders and "normal" limit orders, we also include the number of aggressive limit orders that have *not* been canceled (ALO, ALG) and thus represent the frequency of quote updating by low frequency traders. Finally, OPN and CLS are dummy variables representing the opening and closure period. To be able to aggregate estimates across the market, all variables (expect for dummies, i.e., TYP, OPN and CLS) are normalized to have zero mean and unit standard deviation.

3.4 Econometric Modelling

The chosen categorizations straightforwardly motivate modeling hidden order locations based on an ordered response model. This has several advantages: Firstly, censored observations are straightforwardly taken into account. Secondly, relating market condition variables (as constructed in the previous section) to order categories rather than to plain distances s and d, requires imposing less assumptions on functional form (e.g., linearity) and allows reducing the impact of extreme observations (e.g., executions against hidden depth deep in the book). Thirdly, given the high number of observations (combined with a significant cross-sectional dimension), a reduction of the computational burden is crucial to make the approach tractable. In fact, exploiting the Gaussianity and global concavity of objective functions in an ordered probit model allows to significantly reduce computation time in contrast to, for instance, a (censored) count data model (e.g., negative binomial model) for the variables s or d.

Therefore, we propose modelling hidden order placements using a censored ordered probit model. In order to test our hypotheses, it is sufficient to utilize only order messages which provide information (censored or non-censored) on the location of undisclosed volume. Consequently, the model is not estimated based on the continuous time series of *all* order book messages but only based on the observations revealing hidden order information. In this sense, we do not require a dynamic (e.g., autoregressive) approach as all information on the current and prevailing state of the market is captured by corresponding regressors.

3.4.1 An Ordered Probit Model with Censoring

Let y_t denote the discrete ordered label representing the underlying categories of undisclosed order placements as described in Section 3.3.4. It is driven by a continuous latent variable y_t^* with the link function given by

$$y_t = \begin{cases} 1, & \text{if } y_t^* \leq \gamma_1 \\ 2, & \text{if } \gamma_1 < y_t^* \leq \gamma_2 \\ \vdots & \\ J-1, & \text{if } \gamma_{J-2} < y_t^* \leq \gamma_{J-1} \\ J, & \text{if } \gamma_{J-1} < y_t^* \end{cases} \tag{3.1}$$

where J is the number of categories and γ_j, $j = 1, \ldots, J-1$ denote unknown thresholds. Furthermore, y_t^* is given by

$$y_t^* = \boldsymbol{\beta}'\boldsymbol{x}_t + \varepsilon_t \tag{3.2}$$

with $\boldsymbol{x}_t$ being a $(K \times 1)$ vector of regressors as defined in Section 3.3.5, $\boldsymbol{\beta}$ is a vector of unknown parameters and ε_t denotes an i.i.d. standard normally distributed variable. If the response variable y_t is observed (i.e., non-censoring), the likelihood function is given by

$$L_t^{\mathrm{U}} = \begin{cases} \Phi(\gamma_1 - \boldsymbol{\beta}'\boldsymbol{x}_t) & \text{if } y_t = 1, \\ \Phi(\gamma_j - \boldsymbol{\beta}'\boldsymbol{x}_t) - \Phi(\gamma_{j-1} - \boldsymbol{\beta}'\boldsymbol{x}_t) & \text{if } y_t \in \{2, \ldots, j, \ldots, J-1\}, \\ 1 - \Phi(\gamma_{J-1} - \boldsymbol{\beta}'\boldsymbol{x}_t) & \text{if } y_t = J, \end{cases} \tag{3.3}$$

where $\Phi(\cdot)$ denotes the cumulative density function of the standard normal distribution. In cases, where y_t is not directly observable but only a censored outcome $\widetilde{y}_t$ related to y_t (according to the scenarios described in Figure 3.3 and 3.4) by

$$\widetilde{y}_t = j, \text{ if } y_t \in \{j+1, \ldots, J\}, \tag{3.4}$$

the likelihood function is given by

$$L_t^{\mathrm{C}} = \begin{cases} 1 - \Phi(\gamma_j - \boldsymbol{\beta}'\boldsymbol{x}) & \text{if } \tilde{y}_t = j \text{ and } j = 1, \ldots, J-2, \\ 1 - \Phi(\gamma_{J-1} - \boldsymbol{\beta}'\boldsymbol{x}) & \text{if } \tilde{y}_t \geq J-1. \end{cases} \tag{3.5}$$

Hence, the resulting log likelihood function is given by

$$l = \sum_{t \in \zeta^{\mathrm{U}}} \log(L_t^{\mathrm{U}}) + \sum_{t \in \zeta^{\mathrm{C}}} \log(L_t^{\mathrm{C}}), \tag{3.6}$$

where ζ^{U} and ζ^{C} denote the index sets of uncensored and censored observations, respectively.

The marginal effects of the regressors are straightforwardly given by

$$\begin{aligned} \boldsymbol{q}_1 &= \frac{\partial F_1}{\partial \boldsymbol{x}} = -\phi(\gamma_1 - \boldsymbol{\beta}'\boldsymbol{x})\boldsymbol{\beta}, \\ \boldsymbol{q}_2 &= \frac{\partial F_2}{\partial \boldsymbol{x}} = -(\phi(\gamma_2 - \boldsymbol{\beta}'\boldsymbol{x}) - \phi(\gamma_1 - \boldsymbol{\beta}'\boldsymbol{x}))\boldsymbol{\beta}, \\ &\vdots \\ \boldsymbol{q}_J &= \frac{\partial F_J}{\partial \boldsymbol{x}} = \phi(\gamma_{J-1} - \boldsymbol{\beta}'\boldsymbol{x})\boldsymbol{\beta}, \end{aligned} \tag{3.7}$$

which are commonly evaluated at the sample mean $\bar{\boldsymbol{x}}$.

In case of the dummy variables x_d, we compute the marginal effects as

$$\Delta F_j = \mathbb{P}(y = j | \boldsymbol{x}_{(x_d=1)}, \boldsymbol{\gamma}, \boldsymbol{\beta}) - \mathbb{P}(y = j | \boldsymbol{x}_{(x_d=0)}, \boldsymbol{\gamma}, \boldsymbol{\beta}), \tag{3.8}$$

where $\boldsymbol{x}_{(x_d=i)}$ is a vector with the dummy variable x_d set to i and all other elements being equal to $\boldsymbol{x}$. Appendix B.1 gives the asymptotic distribution of $\widehat{q}_j$ and $\Delta\widehat{F}_j$, the maximum likelihood estimator for q_j and ΔF_j, respectively.

3.4.2 Cross-Sectional Aggregation

We estimate the econometric model on a stock-by-stock basis. For the sake of brevity and compactness of presentation, we aggregate the corresponding estimates across stocks. To explicitly account for differences in estimation precision, we assess the cross-sectional statistical significance relying on a Bayesian framework attributable to DuMouchel (1994) and further implemented by Bessembinder, Panayides, and Venkataraman (2009). Assume that a parameter estimate associated with stock i, $\hat{\beta}_i$, is normally distributed with

$$\widehat{\beta}_i \sim i.i.d.N(\beta_i, s_i^2)$$

and

$$\beta_i \sim i.i.d.N(\beta, \sigma^2),$$

where s_i^2 is the estimated variance of parameter i and variances σ^2 are estimated by maximum likelihood. Then, the aggregated estimate β is obtained by summing up the weighted estimates for all stocks as

$$\widehat{\beta} = \sum_{i=1}^{N} w_i \widehat{\beta}_i, \quad w_i = \frac{(s_i^2 + \hat{\sigma}^2)^{-1}}{\sum_{j=1}^{N}(s_j^2 + \hat{\sigma}^2)^{-1}}. \tag{3.9}$$

Assuming independence across stocks, the variance of the aggregated estimate is given by

$$var(\hat{\beta}) = \frac{1}{\sum_{j=1}^{N}(s_j^2 + \hat{\sigma}^2)^{-1}}.$$

3.5 Empirical Evidence

We estimate separate models for both ask and bid hidden orders for categorizations based on both distance measures s and d. Hence, covering 99 stocks over the cross-section of the market, we estimate 396 models in total. Table 3.5 presents the ordered probit estimates aggregated across all stocks. Recall that lower category labels are associated with a higher hidden order aggressiveness, thus *negative* coefficients reflect that undisclosed orders are set (marginally) *deeper* in the spread. To assess the explanatory power, we report the pseudo-R^2 proposed by McKelvey and Zaviona (1975),

$$R^2_{MZ} = \frac{\sum_{t=1}^{T}(\widehat{y}_t^* - \bar{y}^*)^2}{\sum_{t=1}^{T}(\widehat{y}_t^* - \bar{y}^*)^2 + T} \tag{3.10}$$

where $\widehat{y}_t^*$ is the fitted value of the latent variable y_t^* with ML estimate of $\boldsymbol{\beta}$ and $\bar{y}^* = 1/T\sum_{t=1}^{T}\widehat{y}_t^*$.

Moreover, to provide also insights into the cross-sectional variation of estimates we show histograms of the significant estimates (5% significance level) in Figures B.1 to B.4 in Appendix B.2. Note that the Bayesian cross-sectional aggregates, as discussed in

Table 3.5
Ordered probit estimates
Ordered probit estimates of hidden order locations on the bid and ask side depending on categorized distance measures s and d as discussed in Section 3.3.4. The order aggressiveness is declining with the category label, thus negative coefficients are associated with increasing aggressiveness. Based on 99 NASDAQ stocks during the period of October 2010 with 21 trading days, reported estimates and t-statistics (in parentheses) are cross-sectional aggregates across all stocks using the Bayesian framework of DuMouchel (1994). The reported R^2 is McKelvey and Zaviona's (1975) pseudo R^2.

	Undisclosed bid limit orders				Undisclosed ask limit orders			
	Neg. distance s		Distance d		Neg. distance s		Distance d	
SPR	0.04	(1.1)	1.37	(29.0)	0.06	(1.7)	1.34	(32.2)
DPO	−0.13	(−7.2)	0.01	(2.0)	−0.12	(−7.0)	0.02	(2.5)
DPG	0.00	(0.1)	−0.06	(−6.5)	−0.01	(−0.9)	−0.06	(−6.6)
TYP	0.06	(4.7)	0.10	(7.1)	0.07	(4.9)	0.11	(7.9)
RET	−0.09	(−11.0)	−0.10	(−11.0)	−0.08	(−12.9)	0.10	(−12.2)
VOL	0.01	(1.7)	0.01	(1.9)	0.03	(2.6)	0.02	(2.4)
HVO	−0.40	(−28.3)	−0.36	(−32.9)	−0.43	(−27.4)	−0.39	(−35.7)
HRO	0.17	(14.7)	0.14	(15.7)	0.18	(17.4)	0.16	(16.4)
HVG	−0.01	(−0.9)	−0.02	(−3.2)	−0.00	(−0.3)	−0.01	(−1.1)
HRG	0.03	(4.8)	0.03	(6.6)	0.03	(4.4)	0.03	(4.5)
ALO	0.04	(6.1)	0.01	(2.4)	0.04	(8.4)	0.02	(3.0)
ALG	0.03	(5.2)	0.02	(3.7)	0.04	(7.3)	0.03	(5.0)
HFO	0.07	(6.7)	0.03	(2.6)	0.08	(6.6)	0.03	(2.7)
HFG	0.15	(10.6)	0.11	(11.1)	0.18	(14.2)	0.15	(14.5)
OPN	−0.01	(−0.3)	0.11	(2.8)	−0.04	(−1.0)	0.07	(1.8)
CLS	0.14	(5.7)	0.17	(5.9)	0.13	(5.2)	0.14	(4.7)
Pseudo-R^2	0.29		0.67		0.31		0.68	

Section 3.4.2, are generally close to the averages of significant estimates as these estimates are more precise and have more weight in eq. (3.9). Finally, (Bayesian averaged) estimates of marginal effects for the individual groups of low-, medium- and large-spread stocks are given in Tables 3.6 and 3.7.

In the following, we will discuss the individual results in light of the testable hypotheses formulated in Section 3.2.3. As estimates of parameters and marginal effects are not always straightforward to interpret, we partly illustrate the resulting effects graphically. For the sake of brevity, we will discuss the findings for undisclosed orders on the buy (bid) side only. The corresponding effects on the ask side are strongly symmetric.

3.5.1 Hidden Order Placements in Dependence of Spread Sizes

We find that the size of the bid-ask spread (SPR) has a significant impact on the probability of hidden order placements inside of the spread. The marginal effects associated

Table 3.6

Marginal effects: Aggressiveness of undisclosed orders according to their distance to the own side (distance measure s as shown in Section 3.3.4)

The order aggressiveness is declining with the category label, thus negative coefficients are associated with increasing aggressiveness. Based on 99 NASDAQ stocks during the period of October 2010 with 21 trading days, reported estimates are cross-sectional aggregates using the Bayesian framework of DuMouchel (1994) for the underlying groups of small-, medium- and large-spread stocks. The marginal effects are evaluated at the sample mean. Significant estimates (5% level) are highlighted in the bold font. All values are given in percentages.

	Small spread		Medium spread			Large spread			
	$\mathbb{P}[y=1]$	$\mathbb{P}[y=2]$	$\mathbb{P}[y=1]$	$\mathbb{P}[y=2]$	$\mathbb{P}[y=3]$	$\mathbb{P}[y=1]$	$\mathbb{P}[y=2]$	$\mathbb{P}[y=3]$	$\mathbb{P}[y=4]$
Panel A: Undisclosed buy limit orders									
SPR	**-6.68**	**6.76**	0.22	0.44	-0.69	**0.21**	**4.76**	**2.06**	**-7.30**
DPO	**5.54**	**-5.68**	**0.32**	**0.65**	**-1.02**	**0.10**	**1.39**	**0.50**	**-2.30**
DPG	-0.23	0.46	-0.05	-0.04	0.07	-0.02	**-0.54**	**-0.20**	**0.87**
TYP	**-2.10**	**2.24**	**-0.15**	**-0.38**	**0.60**	**0.06**	**1.01**	0.15	**-1.40**
RET	**0.66**	**-0.78**	**0.37**	**0.76**	**-1.17**	**0.11**	**1.82**	**0.64**	**-2.86**
VOL	0.04	-0.03	-0.07	-0.19	0.32	-0.01	-0.32	-0.02	0.37
HVO	**5.41**	**-5.64**	**2.02**	**4.04**	**-6.28**	**0.39**	**5.97**	**2.96**	**-9.63**
HRO	**-1.04**	**1.10**	**-0.82**	**-1.76**	**2.75**	**-0.15**	**-2.47**	**-1.41**	**4.35**
HVG	**-0.52**	**0.80**	0.06	**0.21**	**-0.32**	0.04	1.11	0.26	**-1.50**
HRG	-0.13	0.13	**-0.09**	**-0.20**	**0.34**	-0.07	**-1.00**	**-0.27**	**1.54**
ALO	-0.17	0.18	**-0.14**	**-0.32**	**0.51**	-0.04	**-0.52**	-0.05	**0.94**
ALG	**-0.64**	**0.69**	**-0.14**	**-0.29**	**0.45**	-0.01	-0.14	-0.05	0.24
HFO	**-0.67**	**0.69**	**-0.22**	**-0.61**	**1.01**	-0.04	-0.74	-0.16	1.15
HFG	**-0.74**	**0.83**	**-0.73**	**-1.55**	**2.39**	**-0.20**	**-3.09**	**-1.37**	**5.09**
OPN	0.51	**-0.63**	0.01	0.11	-0.21	0.02	0.17	-1.43	1.44
CLS	0.13	-0.19	**-0.86**	**-1.79**	**2.79**	**-0.09**	**-2.22**	**-1.89**	**4.91**
Panel B: Undisclosed sell limit orders									
SPR	**-6.01**	**6.05**	0.09	0.40	-0.49	**0.25**	**4.28**	**1.07**	**-5.76**
DPO	**5.34**	**-5.45**	**0.35**	**0.75**	**-1.20**	**0.10**	**1.45**	**0.54**	**-2.26**
DPG	-0.74	0.76	0.06	0.20	-0.31	**-0.04**	**-0.61**	-0.14	**0.87**
TYP	**-2.63**	**2.92**	**-0.23**	**-0.52**	**0.95**	**0.10**	**1.28**	**0.20**	**-1.66**
RET	**0.50**	**-0.52**	**0.39**	**0.82**	**-1.25**	**0.11**	**1.58**	**0.47**	**-2.39**
VOL	-0.22	0.24	-0.13	-0.31	0.49	-0.00	-0.11	0.07	0.04
HVO	**5.68**	**-5.86**	**2.49**	**4.78**	**-7.47**	**0.44**	**6.61**	**2.92**	**-10.26**
HRO	**-1.19**	**1.21**	**-0.98**	**-1.95**	**3.13**	**-0.17**	**-2.84**	**-1.52**	**4.72**
HVG	**-0.28**	**0.30**	0.02	0.04	-0.06	0.05	0.91	0.15	**-1.25**
HRG	-0.11	0.11	**-0.14**	**-0.31**	**0.46**	**-0.05**	**-0.79**	-0.15	**1.14**
ALO	-0.17	0.17	**-0.25**	**-0.49**	**0.81**	**-0.01**	**-0.66**	**-0.26**	**1.13**
ALG	**-0.59**	**0.65**	**-0.05**	**-0.21**	**0.30**	-0.00	**-0.49**	**-0.15**	**0.77**
HFO	**-0.66**	**0.68**	**-0.20**	**-0.52**	**0.81**	-0.04	**-1.31**	**-0.54**	**2.29**
HFG	**-1.04**	**1.09**	**-0.97**	**-2.01**	**3.11**	**-0.24**	**-3.49**	**-1.49**	**5.57**
OPN	**0.82**	**-0.86**	-0.52	-0.84	1.02	0.10	1.87	0.60	-2.93
CLS	0.34	-0.41	**-1.01**	**-2.13**	**3.35**	**-0.13**	**-2.24**	**-0.95**	**3.76**

Table 3.7

Marginal effects: Aggressiveness of undisclosed orders according to their distance to the opposite side (distance measure d as shown in Section 3.3.4)

The order aggressiveness is declining with the category label, thus negative coefficients are associated with increasing aggressiveness. Based on 99 NASDAQ stocks during the period of October 2010 with 21 trading days, reported estimates are cross-sectional aggregates using the Bayesian framework of DuMouchel (1994) for the underlying groups of small-, medium- and large-spread stocks. The marginal effects are evaluated at the sample mean. Significant estimates (5% level) are highlighted in boldfat. All values are given in percentages.

	Small spread		Medium spread			Large spread			
	$\mathbb{P}[y=1]$	$\mathbb{P}[y=2]$	$\mathbb{P}[y=1]$	$\mathbb{P}[y=2]$	$\mathbb{P}[y=3]$	$\mathbb{P}[y=1]$	$\mathbb{P}[y=2]$	$\mathbb{P}[y=3]$	$\mathbb{P}[y=4]$
Panel A: Undisclosed buy limit orders									
SPR	**-27.33**	**28.35**	**-4.10**	**-8.93**	**13.46**	**-0.20**	**-5.06**	**-5.55**	**11.42**
DPO	**-0.11**	**0.11**	-0.01	-0.04	0.06	-0.00	0.00	0.00	-0.00
DPG	**1.08**	**-1.22**	**0.10**	**0.20**	**-0.33**	0.00	-0.02	-0.03	0.04
TYP	**-1.47**	**4.34**	**-0.25**	**-0.54**	**0.85**	-0.01	0.02	0.12	-0.12
RET	0.05	-0.07	**0.29**	**0.64**	**-0.98**	0.06	**0.59**	**0.68**	**-1.35**
VOL	0.09	-0.09	-0.01	0.01	-0.02	-0.01	-0.00	-0.00	0.00
HVO	**6.88**	**-6.89**	**1.26**	**2.58**	**-3.92**	**0.08**	**1.26**	**1.45**	**-2.92**
HRO	**-0.38**	**0.51**	**-0.44**	**-0.92**	**1.46**	-0.00	**-0.38**	**-0.46**	**1.04**
HVG	0.03	-0.04	0.05	**0.16**	**-0.24**	0.02	**0.15**	**0.21**	**-0.42**
HRG	-0.06	0.06	**-0.07**	**-0.16**	**0.25**	-0.01	**-0.08**	**-0.14**	**0.28**
ALO	-0.00	0.02	-0.05	-0.03	0.05	-0.00	0.00	-0.01	0.02
ALG	**-0.26**	**0.33**	**-0.03**	**-0.12**	**0.16**	0.01	0.03	0.04	-0.10
HFO	**-0.18**	**0.19**	-0.04	**-0.15**	**0.22**	-0.00	-0.00	-0.00	-0.27
HFG	**-0.28**	**0.31**	**-0.38**	**-0.85**	**1.32**	-0.06	**-0.27**	**-0.39**	**0.84**
OPN	**0.33**	**-0.37**	0.05	0.42	-0.73	**-0.08**	**-1.79**	**-1.78**	**3.91**
CLS	0.09	-0.10	**-0.26**	**-0.60**	**0.98**	**-0.02**	**-1.22**	**-1.36**	**2.82**
Panel B: Undisclosed sell limit orders									
SPR	**-25.00**	**27.46**	**-4.44**	**-8.56**	**13.55**	**-0.31**	**-5.67**	**-6.28**	**13.07**
DPO	**-0.27**	**0.27**	-0.01	-0.06	0.09	0.00	-0.01	-0.04	0.04
DPG	**1.02**	**-1.11**	**0.15**	**0.25**	**-0.41**	-0.00	-0.03	-0.05	0.05
TYP	**-1.78**	**2.34**	**-0.27**	**-0.62**	**1.07**	0.00	0.06	0.06	-0.08
RET	**0.29**	**-0.31**	**0.25**	**0.60**	**-0.92**	**0.01**	**0.57**	**0.62**	**-1.30**
VOL	-0.03	0.08	-0.01	-0.07	0.09	0.00	-0.00	-0.00	0.01
HVO	**6.38**	**-6.71**	**1.61**	**2.81**	**-4.61**	**0.03**	**1.64**	**1.68**	**-3.74**
HRO	**-0.81**	**0.85**	**-0.49**	**-1.02**	**1.66**	**-0.00**	**-0.45**	**-0.55**	**1.26**
HVG	-0.03	0.03	0.02	0.05	-0.07	-0.00	0.09	0.13	-0.22
HRG	-0.05	0.06	**-0.06**	**-0.16**	**0.23**	-0.02	-0.11	-0.16	0.30
ALO	-0.02	0.02	-0.01	-0.07	**0.11**	-0.00	-0.04	-0.08	0.15
ALG	**-0.24**	**0.28**	-0.01	**-0.10**	**0.14**	-0.00	-0.01	-0.03	0.05
HFO	**-0.31**	**0.33**	-0.05	-0.15	**0.21**	0.00	0.06	0.03	-0.13
HFG	**-0.58**	**0.64**	**-0.60**	**-1.11**	**1.82**	**-0.00**	**-0.60**	**-0.64**	**1.50**
OPN	**0.49**	**-0.66**	0.27	0.58	-1.05	**-0.02**	-0.63	-0.82	1.47
CLS	0.27	-0.30	-0.21	**-0.44**	0.74	**-0.02**	**-1.02**	**-1.19**	**2.55**

with the hidden order distance to the own side shown in Table 3.6 indicate that the effect is negative for small-spread stocks but is positive for large-spread stocks.[4] In particular, for small spread stocks, one standard deviation increase of the spread implies a decrease of the probability of hidden depth inside of the spread (category $y = 1$) by approximately 6.7%. This finding supports the notion of liquidity-induced order placement which tends to be reduced if the spread widens and thus uncertainty rises (Hypothesis (1.A)). However, we find converse evidence for large-spread stocks where the probability of hidden depth inside of the spread increases by approximately 7.3% if the spread widens by one standard deviation. This in line with Hypothesis (1.B) suggesting that hidden order placements in (wider) spreads tends to be rather information-driven than liquidity-driven.

Hence, we show that hidden order strategies tend to be different in low-spread stocks compared to those in large-spread stocks. This might be explained by different motivations and types of market participants driving trading in small-spread vs. large-spread stocks. In particular, we conclude from our findings that liquidity traders are more likely to use aggressive hidden orders in small-spread stocks, but informed traders in high-spread stocks. We explain this by two reasonings. Firstly, small-spread stocks are generally more liquid. This makes it harder to hold informational advantage over a longer time period. Consequently, informed traders prefer visible orders or reserve orders which have time priority compared to hidden orders. In large-spread stocks, however, there is more room to exploit and camouflage informational advantage which make hidden orders more attractive than visible orders. Secondly, in small-spread stocks, the grid of available tick points available for improving the quote is typically too limited. Therefore, even the most passive hidden order inside of the spread can be still too aggressive for an informed traders.

Measuring the aggressiveness of hidden depth in terms of the distance to the opposite side of the market (measure d), we find an overall negative effect. Hence, as shown by the estimates in Tables 3.5 and 3.7, an undisclosed order is placed further away from the opposite side when the spread widens. This effect is consistent with the findings for small-spread stocks based on measure s, but it seems to be a contradiction of the results for large-spread stocks. However, as the categorizations of the two distance measures are partially overlapping, this seeming contradiction is explained by an increasing clustering of aggressive hidden orders in regions close to the own side. The underlying mechanism is graphically illustrated in Figure 3.7 showing how the hidden depth concentrates more significantly on the own side if the spread widens. Intuitively, (informed) traders use hidden orders to compete for the provision of liquidity with own-side liquidity suppliers (and thus increase execution probabilities) while still balancing picking-off risks by remaining sufficiently "passive". As a consequence of symmetric effects for sell orders, we can conclude that the "hidden" spread is positively correlated with the observable spread.

[4] As the estimates reported in Table 3.5 are aggregates across all stocks, they turn out to be insignificant.

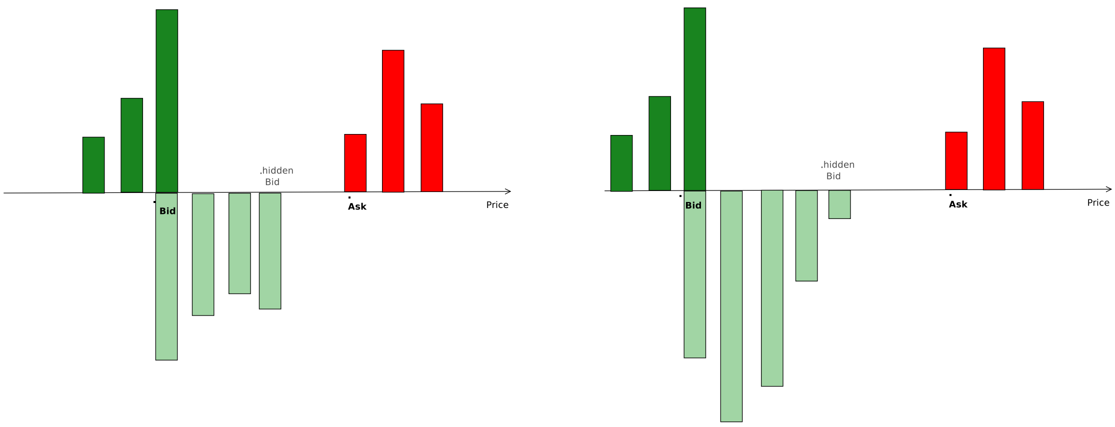


Figure 3.7: Stylized illustration of the effect of a widening of bid-ask spreads on hidden order placements for large-spread stocks. **Left:** scenario of a narrow spread; **right:** scenario of a wide spread. This illustration shows the effect of an increasing hidden order aggressiveness in terms of the distance to the own-side quote, coming along with a decreasing hidden order aggressiveness in terms of the distance to the opposite-side quote.

3.5.2 (How) Does Hidden Liquidity Compete with Visible Liquidity Provision?

Analyzing the effects of visible on-side depth on the best quotes (DPO), we find a clear confirmation of Hypothesis (2.A). Hence, the probability of hidden depth inside of the spread is positively related to the own side visible depth. Table 3.6 reports that the probability of using aggressive hidden bid limit orders increases by approximately 5.5% (2.3%) as the visible depth at the best bid increases by one standard deviation. Hence, traders increase the aggressiveness of hidden orders in order to compete for the provision of liquidity and thus to increase execution probabilities. We again observe contradicting effects for the two underlying distance measures in the case of small-spread stocks leading to the same effects as illustrated in Figure 3.7. Hence, in case of small-spread stocks offering not much room for (hidden) quote improvements, a higher (visible) own-side depth leads to a stronger clustering of hidden orders close to the own-side quote.

While Hypothesis (2.A) is clearly confirmed, we do not find supporting evidence for Hypothesis (2.B). Indeed, Table 3.6 reports that the distribution of hidden depth tends to move to the own side as the visible depth on the opposite side (DPG) increases in large-spread stocks, while no significant change takes place in small-spread stocks. This evidence is in line with the notion that hidden liquidity suppliers tend to reduce adverse selection risk if the price pressure on the opposite side becomes too high. This effect obviously overcompensates traders' motivation to increase execution risk by becoming more aggressive.

Measuring the aggressiveness of hidden depth in terms of the distance to the opposite-

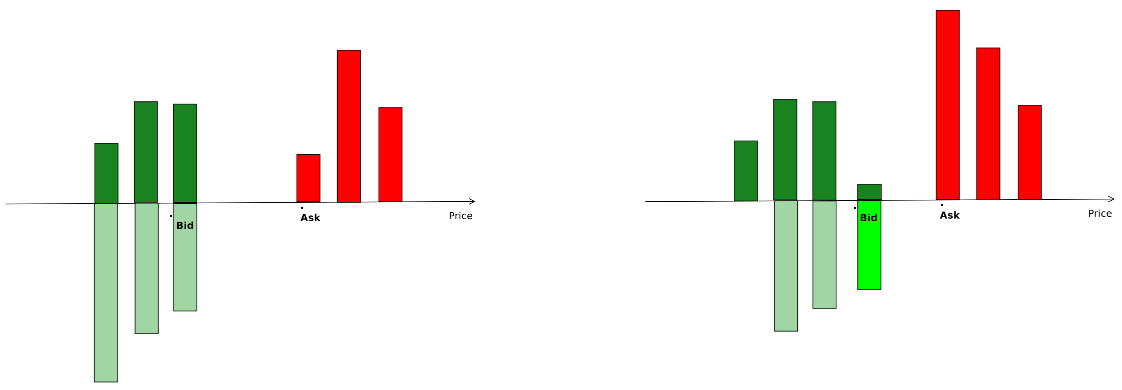


Figure 3.8: Stylized illustration of the effect of an increase of visible ask depth on undisclosed buy order placements for medium-spread stocks. **Left:** scenario of a less visible depth ; **right:** scenario of a huge visible depth. This illustration shows the effect of an increasing hidden order aggressiveness in terms of the distance to the opposite-side quote, coming along with no significant effects on the hidden order aggressiveness in terms of the distance to the own-side quote.

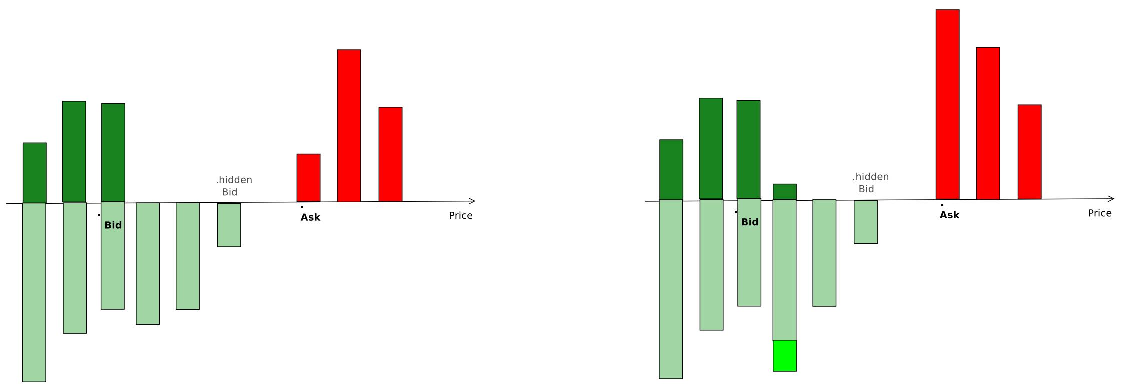


Figure 3.9: Stylized illustration of the effect of an increase of visible ask depth on undisclosed buy order placements for large-spread stocks. **Left:** scenario of a less visible depth ; **right:** scenario of a huge visible depth. This illustration shows the effect of an decreasing hidden order aggressiveness in terms of the distance to the opposite-side quote, coming along with no significant effects on the hidden order aggressiveness in terms of the distance to the opposite-side quote.

side quote (measure d), we find positive effects in small- and medium-spread stocks. As shown by the estimates in Tables 3.5 and 3.7, an undisclosed order is placed closer to the opposite side when the opposite side visible depth increases. The seeming contradiction to estimates in Table 3.6 implying no significant effects could be explained by the increasing use of reserve orders and the partially overlapping categories. We graphically illustrate the underlying mechanism in Figure 3.8 showing how traders update the own-side quote by the reserve order. A similar explanation can be reasoned for the seeming contradiction of estimates in large-spread stocks as well (Figure 3.9). The finding supports Buti and Rindi (2011)'s theoretical prediction that the use of reserve orders, rather than hidden orders, increases with the opposite-side visible depth. Pardo Tornero and Pascual (2007) and De Winne and D'Hondt (2007) find similar evidence for the Spanish Stock Exchange and Euronext Paris where only reserve orders but not hidden orders can be used.

3.5.3 Hidden Order Placements After Price Movements and Trading Signals

We show that the aggressiveness of hidden bid depth decreases when the prevailing trade is seller-initiated (TYP). In particular, the hidden bid depth shifts away from the ask side. This reduces liquidity suppliers' risk of being picked off by (eventually better informed) sellers but increases their risk of non-execution. Conversely, in case of a buy market order, hidden liquidity supply on the bid side increases and moves toward the ask side. Hence, liquidity suppliers follow trading directions in the sense that they post more aggressively and thus increase execution probabilities without facing too high adverse selection risk (as long as buy pressure dominates).

In this sense, Hypothesis (3.A) is confirmed. Besides economic reasoning, a pure mechanical effect may further drive the results. In particular, as a sell trade itself absorbs pending aggressive undisclosed buy limit orders, the aggressiveness of hidden bid depth temporarily decreases. This effect, however, is only true in case of trades arriving instantaneously before the observation of interest. But as our estimates utilize all order messages revealing information on hidden orders (occurring on average 30 times more frequently than trades), these mechanical effects apply only infrequently.

Analyzing the effects of recent price movements (RET) on hidden order placements, we find similar effects and supportive evidence in favor of Hypothesis (3.B). Accordingly, the aggressiveness of undisclosed bid orders increases as prices have been moved upwards. Specifically, the probability of hidden orders inside of the spread increases approximately 2.9% when the return increases by one standard deviation. Moreover, the estimates in Table 3.5 show that hidden bid depth moves closer to the ask side. Again, this supports liquidity suppliers' motivation to reduce the risk of non-execution. Conversely, in case of prevailing negative price movements, hidden liquidity placements on the bid side become less aggressive with the hidden depth distribution shifting away from the ask side. As postulated in Section 3.2.3, this is explained by protection against picking-off risks in case prices continue moving downwards.

Interestingly, no clear confirmation of Hypothesis (4) is found. We do not find significant impacts of prevailing return volatility. According to our estimates, hidden order aggressiveness even tends to increase in volatile market periods. However, in most cases, these effects are insignificant.

3.5.4 Competition for Hidden Liquidity Provision

Our estimates show clear evidence for hidden liquidity competition. According to the estimates associated with the effects of own-side hidden liquidity supply (HVO), we support Hypothesis (5). In particular, Table 3.6 reports that the probability of hidden bid depth inside of the spread increases by approximately 10% as the execution of hidden bid volume during the last minute increases by one standard deviation. This effect is supported by the estimates in Table 3.7 indicating that hidden liquidity shifts closer to the opposite side of the market. Hence, according to the reasoning motivating Hypothesis (5), liquidity suppliers provide further hidden volume if they realize liquidity demand from the opposite side and competition on their own side.

Our estimates show that these effects prevail as long as adverse selection risk does not become too high. Indeed, when the prevailing one minute hidden depth execution becomes too high compared to the corresponding activities during the last five minutes (HRO), hidden order aggressiveness tends to decline. In this situation, price pressure from the opposite side becomes too strong and makes adverse selection risk too high.

Studying the effect of hidden order detections on the opposite side of the market, we find slight evidence for the effect that trading against hidden sell orders also increases the hidden order aggressiveness on the bid side. This might be explained by the fact that buy market orders make buy hidden orders (relatively) less aggressive and move away hidden ask quotes. This, in turn, gives hidden liquidity suppliers on the bid side more room for quote improvements and thus the reduction of non-execution risks.

3.5.5 Hidden Order Placements and HFT

Analyzing the effects of HFT (approximated by the intensity of fleeting orders) on hidden order submissions, we find strong empirical support for Hypothesis (6). Indeed, the more opposite-side traders try to detect hidden liquidity by "pinging activities" (HFG), the lower the hidden order aggressiveness. Especially for large-spread stocks, an one standard deviation increase of HFT activities on the ask side implies a decrease of the probability of hidden bid depth placements inside of the spread by more than 5%. Consequently, the distribution of hidden depth moves away from the opposite quote. Hence, traders interpret the rapid cancellations of limit orders as signals for hidden liquidity detection strategies rather than true liquidity supply. These results are in line with empirical evidence reported by Hasbrouck and Saar (2009) and the predictions by Buti and Rindi (2011) showing that hidden order placements become non-attractive if hidden depth is easily detected.

Note that the effects on the distribution of the entire hidden depth (as reported in

Table 3.7) are substantially smaller than those on hidden depth inside of the spread (as revealed by Table 3.6). This finding also supports the theoretical prediction by Buti and Rindi (2011) that reserve orders, rather than hidden orders, are dominantly used when parasitic traders utilize front running strategies.

3.5.6 Intraday Patterns

We find no clear confirmation of Hypothesis (7.A) postulating a higher hidden order aggressiveness during or after the opening period. Actually, our findings for small-spread stocks support the hypothesis, while it is rejected for large-spread stocks. However, clear evidence for Hypothesis (7.B) is shown. Indeed, for large-spread stocks, we find that the probability for hidden bid depth placements within the spread in the hour before market closure is approximately 5% lower than during the rest of the day. This supports the economic reasoning that displayed orders are preferred if the time horizon becomes shorter and the importance of time priority rises.

3.6 Conclusion

Many stock exchanges around the world choose to reduce market transparency by allowing traders to hide a portion of their order size. As a consequence trading under limited pre-trade transparency becomes increasingly popular in financial markets. Previous studies in the literature examine opaque markets with only partially undisclosed orders. This study sheds light on traders' use of completely undisclosed orders in electronic trading, based on a sample of 99 stocks traded on NASDAQ during October 2010.

Employing NASDAQ TotalView message data, we retrieve information on hidden depths from the visible order activities and propose an ordered response approach with censoring mechanism for modelling the hidden order locations conditional on the state of market. Our finding shows that the hidden liquidity supply has significantly correlation to the market conditions and thus is predictable in terms of the state of the prevailing (visible) LOB and order flow. The evidence suggests that traders make their undisclosed order submission strategies by balancing the non-execution risk and picking-off risk.

Our findings are of interest to academics and institutional trading desks. A better understanding of traders' decisions on using undisclosed orders could help theorists in developing comprehensive models on trader behaviour in opaque markets with both hidden order types. Moreover, these results provide useful insights for institutional traders who are obligated to acquire a big position. They may benefit by either improving their undisclosed order submission strategies or searching for hidden depths on the opposite side of the market more efficiently.

Chapter 4

Extracting Information from the Message Stream

This chapter is based on Huang and Polak (2011).

4.1 Introduction

An electronic limit order market is an order-driven market which automatically collects orders from traders in a centralized limit order book (LOB) and matches corresponding buy and sell orders based on specific priority rules, very often the *price-time priority rule*. Currently, most equity exchanges around the world are either pure electronic limit order markets, e.g. NYSE Arca, BATS, Euronext, Australian Stock Exchange (ASX) and Direct Edge, or at least allow for customer limit orders in addition to on-exchange market making, e.g. NASDAQ, NYSE and the London Stock Exchange (LSE). The traditional monopolistic power of market makers in the area of liquidity provision through quoting on both sides of the market has been strongly restricted, if not completely eliminated. Instead, the important task of providing liquidity is now assigned to the complex trading interactions enabled by the emergence and disclosure of the LOB. Hence, the state of the LOB is extremely important for practitioners, because it allows them to optimize their trading strategies, but also for researchers who analyze trading activity in these markets and try to interpret the underlying economic motivation.

One of the most prominent market structure developments in recent years is high frequency ("HF") trading. HF traders in general employ extremely quick and sophisticated computer programs for generating, routing and executing orders. They establish and liquidate positions in very short time-frames by submitting numerous orders and cancelling non-executed orders shortly after submission. As a consequence, the trading volume grows, orders shrink in size and the pace of LOB updating is beyond human perception, requiring nanosecond precisions. The volume of information about orders recorded by market organizers therefore dramatically increases.

More than ever before researchers today face the challenge of working with real datasets on micro-structure level of financial markets. They are in general not difficult to obtain, but very difficult to process. One way of going about it is acquiring snapshots of historical LOB data. But this is very impractical, because such datasets are usually very large and contain only incomplete information[1]. The second option is to acquire much better compressed raw message stream data, like TotalView-ITCH and Multicast PITCH data. Because message data record all visible order activities, i.e. limit order submissions, cancellations and executions, they can be used to reconstruct the historical LOB up to any required precision (level). However, this creates a different type of challenge; it is necessary to reconstruct the LOB using the same rules that were used by the matching algorithm applied by the exchange. Although simple in principle, such algorithms need to take into account market-specific issues and, considering the large volume of data, work extremely efficiently.

In this chapter, we present you the LOBSTER, a program framework for reconstructing LOBs as well as extracting order flow information from historical message stream data. Note that the fundamental limit order activities are quite similar across different limit order markets, though the specific trading rules can be very different. We modularize LOBSTER to processes limit order activities translated from messages instead of messages themselves, so that it can be easily adapted to data from new limit order markets with just a few modifications. The underlying data structure in these modules is highly optimized and their programs are exhaustively tested to guarantee the reliability of the output data and the efficiency of the entire system.

Currently we implement a translation module for NASDAQ TotalView-ITCH message and a web-based interface. Researchers around the world can easily access our program through `http://lobster.wiwi.hu-berlin.de` and download NASDAQ LOB data reconstructed on the fly. Moreover, a related forum facilitates general discussion on empirical analysis of LOB, as well as bundles of small programs and useful tools in both R and Matlab.

The remainder of this chapter is structured as follows. In Section 4.2, we take the NASDAQ TotalView-ITCH as an example to illustrate the structure of order-related messages in message stream data. Section 4.3 gives the overview of the design of the LOBSTER. We discuss in detail on two implemented modules, the LOB Constructor and the Order Tracer, in Section 4.4 and 4.5, respectively. Finally, Section 4.6 concludes.

4.2 Message Stream Data

As we already discussed above, unlike historical snapshots of the LOB, message-type data allow reconstructing the full LOB and observing its dynamics with maximum precision. This is indeed extremely valuable for academic research, giving the possibility to study all aspects of trading in full detail. In this section, we will take the NASDAQ TotalView-

[1]Snapshots are made at regular time intervals, e.g. every second, which means that multiple changes within this time intervals may be omitted. Should snapshots capture *all* changes in the LOB to *all* levels the size of files to store these datasets would be too large for practical purposes.

Table 4.1
TotalView-ITCH: order-related messages.
TotalView-ITCH includes messages representing *submissions*, *executions* and *cancellations* of limit orders, as well as *executions of hidden orders*. A *submission* type limit order message may or may not contain information identifying the market participant, who submitted the order - Market Participant Identification (MPID). A type "P" message reports the execution price and immediately traded quantity, i.e. only that part of the hidden order, which is currently being executed against an incoming order. Thus no information about the remaining unexecuted part of the hidden order is revealed.

Instruction	Type	Limit/Hidden Order				Canc./Exe.
		ID	Price	Size	MPID	size
Limit order submission	A	✓	✓	✓		
Limit order submission with MPID	F	✓	✓	✓	✓	
Limit order execution	E	✓				✓
Limit order cancellation (partially)	X	✓				✓
Limit order cancellation (totally)	D	✓				
Hidden order execution	P	✓	✓			✓

ITCH 4.0 as an example to illustrate the structure of messages.

The original NASDAQ TotalView-ITCH is a direct data feed that contains market messages for all *submissions*, *cancellations* and *executions* of limit orders, as well as *executions of hidden orders*. TotalView-ITCH 4.0 data is the binary version of the TotalView-ITCH data introduced in November 2008, following the ITCH 3.0 format. This format is ready for nanosecond time stamps[2] and contains longer order and trade identification numbers allowing to mark up to 18 quadrillion messages per day.

Considering the enormous number of messages generated by the activities in markets every day, the TotalView message stream is designed to present information in a parsimonious way, reducing redundancy in the records. Table 4.1 illustrates the message types related to the instructions carried by orders. Whenever a market participant submits a new order, its details, such as order ID, limit price and size, are recorded in a *submission* message. If the market participant chooses to reveal her identity to other market participants the *submission* message contains also her Market Participant Identification (MPID). All subsequent changes to the submitted limit order are also recorded in the form of messages; messages reporting partial cancellations or executions (partial or total) contain the same order ID as the original *submission* messages and the cancelled or executed quantity. If the limit order is totally cancelled (deleted), TotalView records only the corresponding order ID. Finally, messages reporting executions of hidden orders contain only the price and the executed quantity, but *not* the total size of the originally submitted hidden orders.

[2]The time resolution of trading in NASDAQ is still in milliseconds. Therefore, the time stamps in the current data set have still millisecond precision rather than nanosecond.

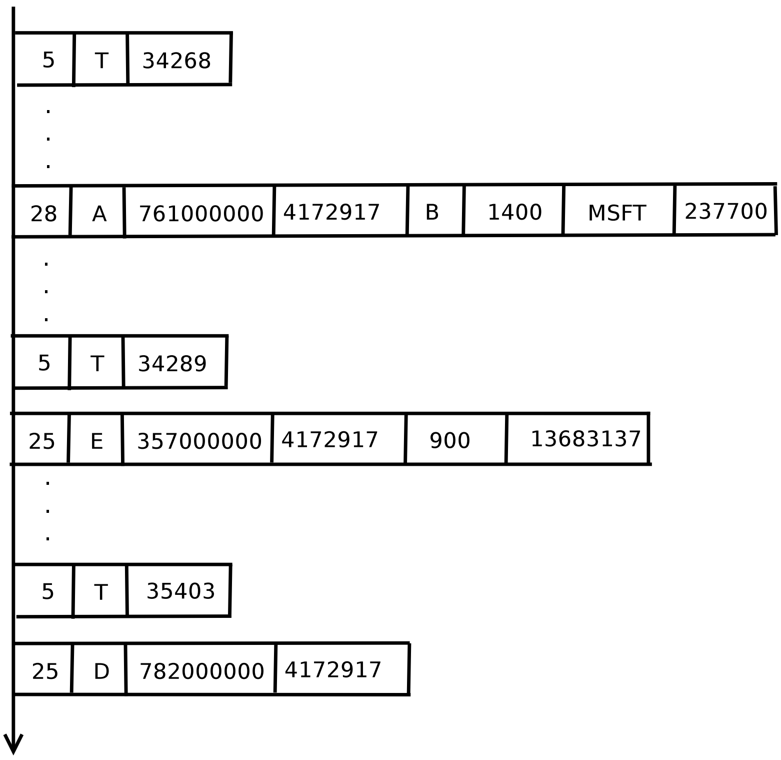
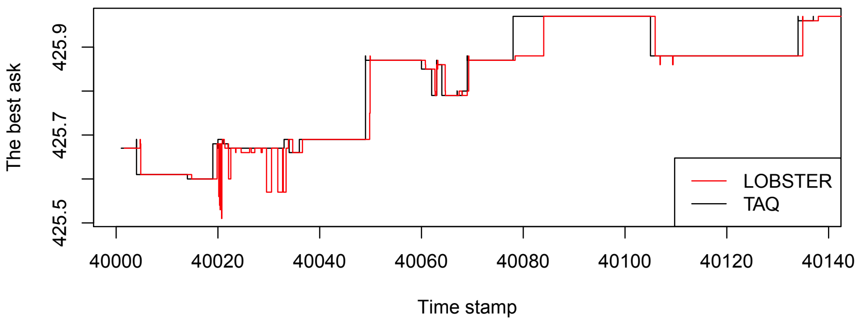


Figure 4.1: A sequence of messages related to the same limit order. The first part of each message contains information about the length of the message. The second part is the message type. Time stamp messages "T" record the number of seconds after midnight. The third part of messages (except for time stamp messages) contains the time in nanoseconds since the last time stamp message. The fourth part is the order ID. For type "A" messages, the remaining fields contain: trade direction (buy/sell) indicator, order size, stock ticker, limit price. Type "E" messages further contain information about the executed size and a unique matching number.



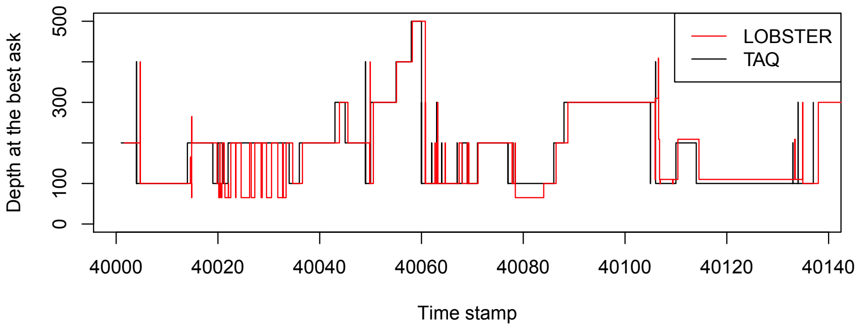


Figure 4.2: Comparison of TAQ and one-level LOB generated by LOBSTER. The time stamp is the number of milliseconds after midnight.

However, this parsimony raises challenges for researchers who intend to use TotalView message data. Most messages contain only incomplete information for the corresponding order. Figure 4.1 illustrates a sequence of messages related to the same limit order. Only the "A" message, which carries information about limit order submission, contains information about limit price, size and trade direction of the limit order. Therefore, before type "E" or "D" messages can be used by the algorithm to update the LOB, we need to "trace back" the corresponding "A" message to retrieve the information. Note also that ITCH records time stamps in two parts. The first part of a time stamp is carried by a type "T" message, which records the number of seconds after midnight. The second part is the number of nanoseconds since the last recorded second, recorded in the third position of each message.

Compared to the Trades and Quotes Database (TAQ) released by NYSE, which contains the best quotes and the corresponding depths, message data has richer information. It records order activities, which pooled together comprise the quote prices and depths.

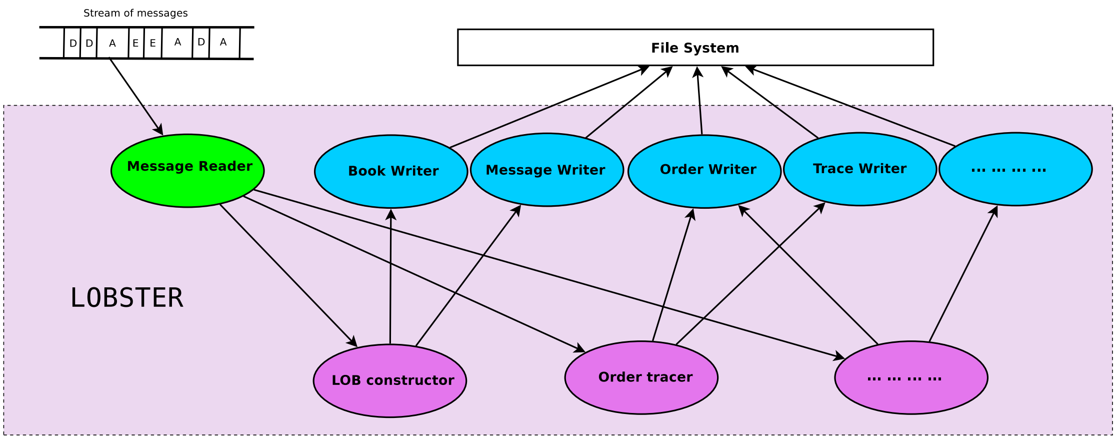


Figure 4.3: The overview of LOBSTER system. Three types of modules are used in general. Readers (green) read the source data into the system; data Processors (purple) retrieve the information; Writers (blue) write the output into the file system. For the sake of flexibility Readers and Writers are normally implemented as interfaces.

Thanks to this superiority of information content message data can be used to reconstruct the LOB up to any quote level. But even when looking only at the best quote and depth, message data is richer than the usual data obtained from the TAQ database. Message data contain information about limit orders which are cancelled shortly after the submission. These limit orders are typically submitted in order to detect hidden liquidity inside the spread rather than to provide liquidity (see e.g., Hasbrouck and Saar, 2009). Therefore, it is not surprising that TAQ ignores them as shown in Figure 4.2. However, these orders are crucial for some studies, e.g the analysis of high frequency trading strategies and hidden order submission strategies.

The currently implemented LOBSTER is connected to a storage facility containing over 5 TB of historical TotalView-ITCH data, in ITCH 3.0 format from the period Jan 2007 to Apr 2009 and ITCH 4.0 format from the period May 2009 to Dec 2010. This dataset contains only limit order messages. Other messages such as imbalance data events and administrative messages have been cleaned out.

4.3 Overview of LOBSTER

The main goal of LOBSTER is to provide a reliable, efficient and flexible platform for researchers to retrieve information from parsimonious message data. Relying on the object-oriented concept, we designed LOBSTER as a modular system with three types of modules: *Readers*(green), *data Processors*(purple) and *Writers*(blue) as shown in Figure 4.3.

Reader translates data from an external source to a stream of order events, which is then processed by data Processors. Based on a unified abstract interface, different readers can be created for reading different input formats, e.g. from different stock exchanges. But also an order-flow simulator can act as a "Reader", provided that the generated data have the required format. Currently the system contains an ITCH message Reader for reading binary ITCH files from the storage facility and a test version PITCH Reader for BATS.

Data Processor is the core of the system responsible for extracting the information required by users from the order flow generated by Reader. Connected to Reader by an abstract interface, the data Processors treat historical order flow, simulated order flow or hybrid order flow identically. This creates great potential for very effective testing of trading strategies.[3]

We implemented a `LOB Constructor` that matches the limit orders coming from the Reader and updates the LOB. The accuracy and efficiency of this algorithm determine to a large extent the overall performance of the system. For this reason it was thoroughly tested and optimized. We shall discuss the currently implemented `LOB Constructor` in Section 4.4.

Moreover, we have also developed `Order Tracer`, whose beta version is currently being tested. `Order Tracer` traces relevant events for individual limit orders, such as the time of submission, partial/full execution and cancellation. It has been used for computing the lifetime of limit orders for some research projects (e.g. Hautsch and Huang, 2012a,c). We expect that in the near future more and more utilities for special research purposes will be added to the current framework.

Writer receives the reconstructed data and saves them to the file system. Currently the system contains four Writers; the `Message Writer`, which saves the event type and the corresponding order information into the file system, the `Book Writer`, which receives the current state of the reconstructed LOB for every order event and saves it, the `Order Writer`, which saves the characteristics (order ID, limit price, size, etc.) of limit orders, and finally the `Trace Writer`, which saves event-specific information, such as the event time, submission time and the type of order event, as generated by the `Order Tracer`.

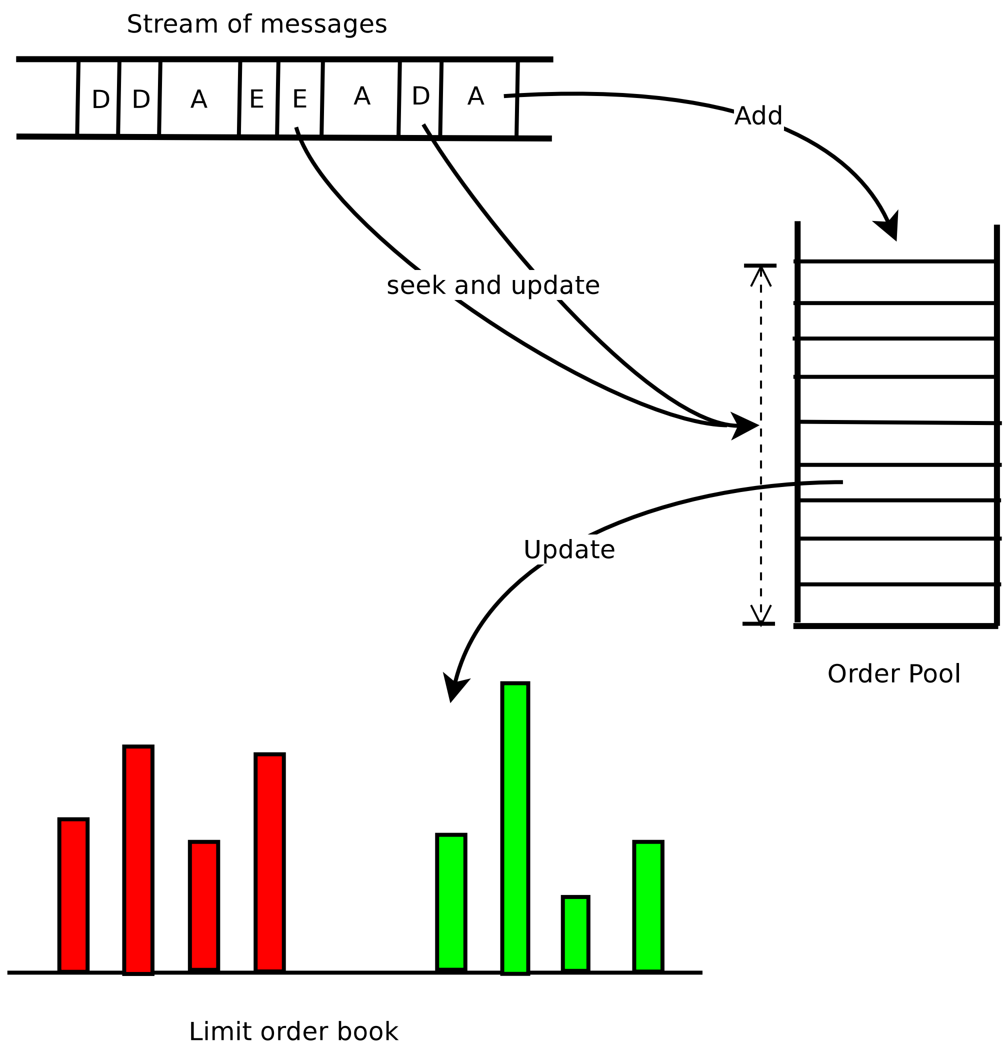


Figure 4.4: LOB reconstruction procedure. The algorithm employs an order pool to collect the limit order information. When an "A" (or "F") message comes in, it creates a limit order item in the order pool. When subsequently a message comes in indicating limit order cancellation ("X" and "D") or a limit order execution ("E"), the information about the price and size of the original limit order is retrieved from the order pool using common order ID.

4.4 Limit Order Book Reconstruction

4.4.1 Overview of the Reconstruction Procedure

Figure 4.4 summarizes the procedure of the LOB reconstruction. For arriving messages identified as limit order submissions, the system records in the order pool all relevant information including order ID, limit price, quantity, trade direction and MPID, if available. Once a cancellation or execution message arrives, the system first finds in the pool the corresponding previously recorded limit order submission by comparing the order IDs. After matching the two orders - incoming order and order stored in the order pool - the system records the remaining non-executed size of the limit order or deletes the order from the order pool altogether if the remaining size is zero. Finally, the system updates the LOB: the side in the LOB (bid or ask) to be updated is determined by trade direction (buy or sell) of the corresponding order; the level in the LOB is identified by the limit price; the new depth at this level is calculated by deducting the size of the effective quantity.

The remaining issue is the construction of the initial-state of the LOB before the aforementioned procedure can be applied. Note that in the TotalView-ITCH message data, the order ID of any limit order cancellation and execution message can always be found in a limit order submission message, which was recorded at an earlier time on the same trading day. This implies that all limit orders valid overnight, such as some good-to-kill orders, have been resubmitted by the system in the early morning. The NASDAQ trading system is in general open for new order instructions at 7:00 EST, even though continuous trading does not start until 9:30 EST. Because the TotalView-ITCH data set contains all messages, including messages submitted during the pre-trading period, our algorithm initializes the reconstruction with an empty LOB at the beginning of every day.

4.4.2 Implementation of LOB Constructor

As we discuss in Section 4.2, all TotalView-ITCH messages, except for messages containing limit order submissions, contain only partial information about the underlying limit orders. In order to update the order book, we need to "complete" the limit order information. Figure 4.5 shows in details how this is done in case of messages containing limit order executions. After reading an execution message, the system searches for the corresponding limit order inside the order pool using the order ID. If the search is successful, the information on the remaining size of the limit order is updated. The system then updates the LOB by changing the depth and quote on the corresponding level.

[3] For instance, it is well-known that the back-testing of trading strategies on historical data has a very serious drawback - there is no market feedback to the tested strategy and so it is very hard to estimate the market impact of the strategy. In this context, testing with a simulator allows implementing realistic feedback mechanisms to simulate market impact and thus provide credible assessment of the tested trading strategy.

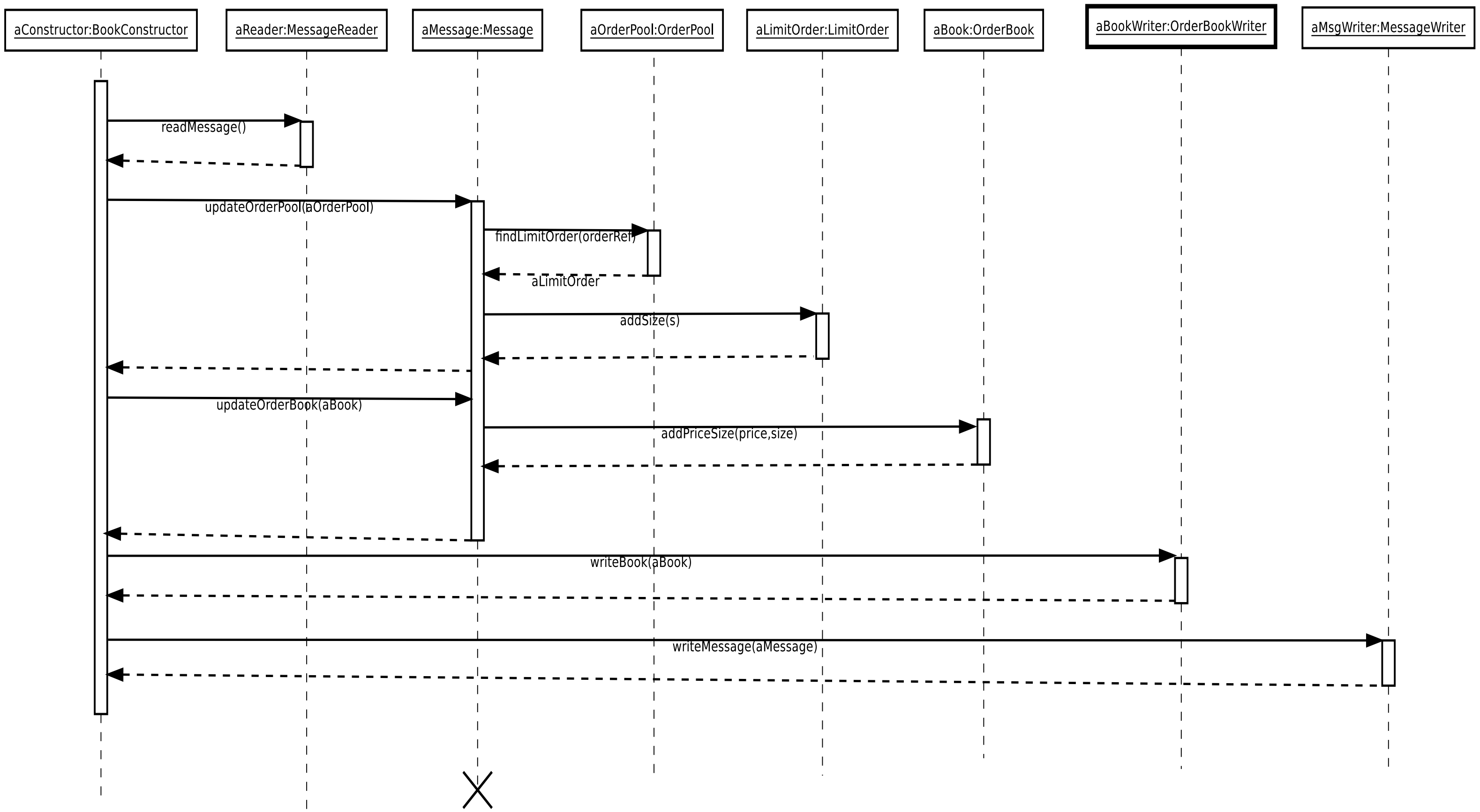


Figure 4.5: Sequential diagram for messages of limit order executions. The procedure begins with reading the message and creating a corresponding `Message` object. The object then communicates with the order pool and updates information about the underlying limit order. Simultaneously, it retrieves from the limit order any missing information, such as price. Using the complete information about the limit order, the `Message` object updates the order book state (the `OrderBook` object). Finally, the constructor writes the updated order book state and the corresponding message into the file system using the `OrderBookWriter` and `MessageWriter` objects.

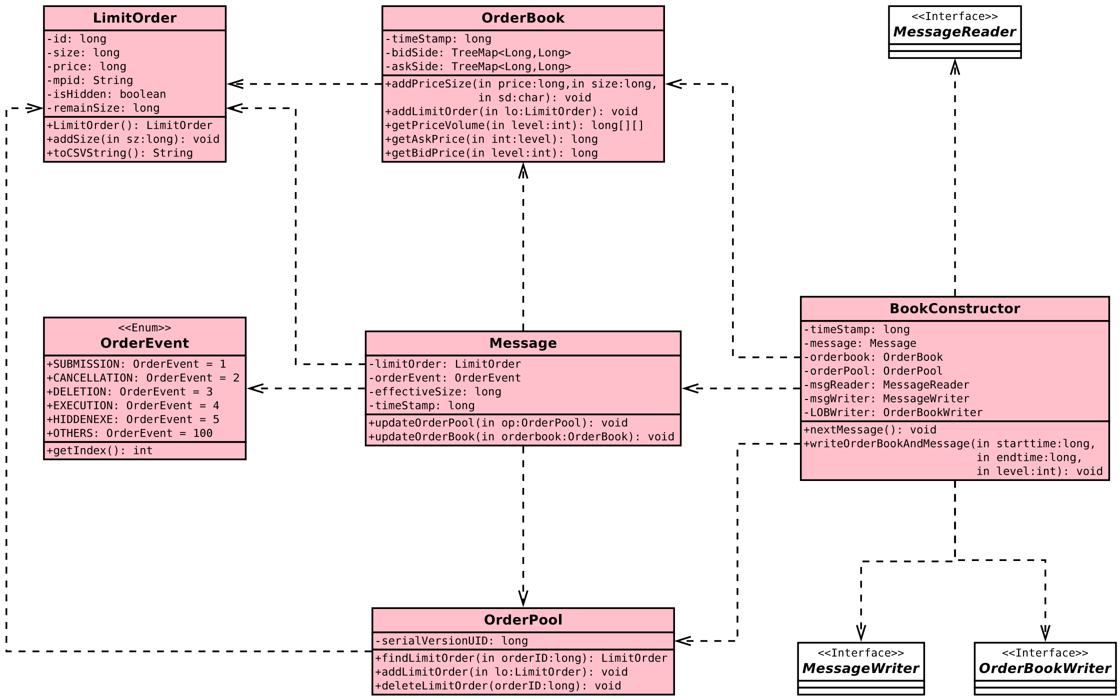


Figure 4.6: Class diagram of LOB construction. A `BookConstructor` object contains unique `OrderBook` and `OrderPool` objects, which are updated using incoming messages corresponding to `Message` objects. Moreover, there must be at least one `MessageReader` object for reading the input, an `OrderBookWriter` object and a `MessageWriter` object, which write the output into the file system.

Finally, the new order book and the corresponding message item are stored in the file system as output.

Figure 4.6 represents the final class diagram for our LOB Constructor. The system includes a unique `OrderBook` and `OrderPool` instance, which are controlled via a unique `BookConstructor` instance. After `MessageReader` objects reads a message from an external source, it creates a `Message` object and then uses it to update the limit orders inside the `OrderPool` object as well as quote and depth information in the `OrderBook` object according to the order event type. If the process is successful, it saves the output using a `MessageWriter` and an `OrderBookWriter` object. Note that the `Reader` and `Writer` are intent to be defined as abstract interfaces, rather than concrete classes, for increasing the flexibility of the format of output data.

4.4.3 Output of LOB Constructor

The `LOB Constructor` generates two output data files; one file contains the LOB data and the other one contains the corresponding order events. Table 4.2 shows a segment of

Table 4.2
LOB data generated by the LOB Constructor.
The sample contains a five second segment of three-level LOB data for ticker GOOG on July 1st 2009. The time variable is in milliseconds after the midnight. Price is in 0.01 of a cent and size in number of shares. The corresponding order events that update the state of the LOB are shown in Table 4.3.

Time	Ask price 1	Ask size 1	Bid price 1	Bid size 1	Ask price 2	Ask size 2	Bid price 2	Bid size 2	Ask price 3	Ask size 3	Bid price 3	Bid size 3
36000043	4231100	100	4227300	300	4231200	100	4223000	100	4231300	300	4222900	100
36000044	4231100	100	4227300	300	4231200	300	4223000	100	4231300	300	4222900	100
36000207	4229100	100	4227300	300	4231100	100	4223000	100	4231200	300	4222900	100
36000208	4229100	100	4227300	300	4231100	100	4223000	100	4231200	100	4222900	100
36000208	4229100	100	4227300	300	4231100	100	4223000	100	4231300	100	4222900	100
36003222	4229100	100	4227300	300	4231100	100	4223000	100	4231200	100	4222900	100
36003471	4229100	100	4227300	300	4231100	100	4222900	100	4231200	100	4221200	100
36004005	4229100	200	4227300	300	4231100	100	4222900	100	4231200	100	4221200	100
36004009	4229100	200	4227300	200	4231100	100	4222900	100	4231200	100	4221200	100
36004009	4229100	200	4222900	100	4231100	100	4221200	100	4231200	100	4219100	400
36004009	4229100	200	4222900	100	4231100	100	4221200	100	4231200	100	4219100	400
36004010	4229100	200	4222900	200	4231100	100	4221200	100	4231200	100	4219100	400
36004010	4229100	200	4227300	200	4231100	100	4222900	200	4231200	100	4221200	100
36004011	4229100	100	4227300	200	4231100	100	4222900	200	4231200	100	4221200	100
36004015	4229100	100	4227300	200	4231100	100	4223300	100	4231200	100	4222900	200
36004016	4229100	100	4227300	200	4231100	100	4222900	200	4231200	100	4221200	100
36004017	4229100	100	4222900	200	4231100	100	4221200	100	4231200	100	4219100	400
36004018	4229100	100	4222900	200	4231100	100	4222800	200	4231200	100	4221200	100
36004018	4229100	200	4222900	200	4231100	100	4222800	200	4231200	100	4221200	100
36004018	4229100	200	4227300	200	4231100	100	4222900	200	4231200	100	4222800	200
36004020	4229100	200	4227300	200	4231100	100	4222900	200	4231200	100	4221200	100
36004020	4229100	100	4227300	200	4231100	100	4222900	200	4231200	100	4221200	100
36004021	4229100	100	4227300	200	4231100	100	4223300	100	4231200	100	4222900	200
36004025	4229100	100	4223300	100	4231100	100	4222900	200	4231200	100	4221200	100
36004025	4229100	100	4222900	200	4231100	100	4221200	100	4231200	100	4219100	400

Table 4.3
Order event data generated by the LOB Constructor.
The sample contains a five second segment of order event (affecting the three-level LOB) data for ticker GOOG on July 1st 2009. The time variable is in milliseconds after the midnight. The type variable indicates the event type: submission, cancellation, execution of a limit order or execution of a hidden orders. Order ID is the ID of the corresponding limit or hidden order. Size is the effective size, i.e. order size for submission events and cancelled (executed) quantity for cancellation (execution) events. The three-level LOB instances immediately after these order events are shown in Table 4.2.

Time	Type	Order ID	Size	Price	Trade Direction
36000043	1	35859474	100	4231100	-1
36000044	1	35859503	200	4231200	-1
36000207	1	35862501	100	4229100	-1
36000208	3	35859503	200	4231200	-1
36000208	3	35603811	100	4231200	-1
36003222	1	35926475	100	4231200	-1
36003471	3	35293758	100	4223000	1
36004005	1	35948533	100	4229100	-1
36004009	4	35332615	100	4227300	1
36004009	4	35643198	200	4227300	1
36004009	5	35643169	200	4227300	1
36004010	1	35948820	100	4222900	1
36004010	1	35948851	200	4227300	1
36004011	3	35948533	100	4229100	-1
36004015	1	35949144	100	4223300	1
36004016	3	35949144	100	4223300	1
36004017	4	35948851	200	4227300	1
36004018	1	35949411	200	4222800	1
36004018	1	35949425	100	4229100	-1
36004018	1	35949469	200	4227300	1
36004020	3	35949411	200	4222800	1
36004020	3	35949425	100	4229100	-1
36004021	1	35949745	100	4223300	1
36004025	4	35949469	200	4227300	1

reconstructed three-level LOB data for Google Inc. (with ticker GOOG in NASDAQ). It includes quotes and the corresponding depths up to the third best ask and bid, together with time stamps. The other file contains the corresponding order event. As shown in Table 4.3, it has five fields:

- `time`: time stamp; milliseconds after mid-night.
- `type`: event type; 1 for limit order submission, 2 for partial cancellation, 3 for total deletion, 4 for limit order execution, 5 for hidden order execution.
- `order ID`: a unique number assigned by the exchange for identification of orders.
- `size`: change of order size (in shares); for limit order submission it is the order size, for order execution it is the trading volume and for limit order cancellation it is the cancelled volume.
- `price`: the price of the limit order or the corresponding executed hidden order (in 0.01 of a cent).
- `trade direction`: the trade direction of the corresponding limit (hidden) order; 1 for buy limit order and −1 for sell limit order. Note that in case of order execution, 1 corresponds to seller-initiated trade (i.e. execution against buy limit order) while −1 corresponds to buyer-initiated trade (i.e. execution against sell limit order).

Both files can be easily loaded and used by analytical software, such as Matlab and R. Because the number of output order events is identical to the number rows in the LOB (i.e. both files have the same number of rows), the two output files can be easily merged. The following is an example of R code for loading and merging the data set.

An Example of R Code Loading LOB Data

```r
######### load data ########################
# load order book data
dataOB <- read.csv("GOOG_20090701_orderbook_3.csv")
# load message data
dataM <- read.csv("GOOG_20090701_message_3.csv")
# merge two datasets
data <- cbind(dataM,dataOB[,-1])
####### load completed ##################

# compute the number of order book levels
nlevels <- (dim(dataOB)[2] - 1)/4
# name the columns
colnms <- c("Time","Type","OrderID","Size","Price","TradeDirection")
for (i in 1:nlevels)
{ colnms <- c(colnms,paste("ASKp",i,sep=""),paste("ASKv",i,sep=""),
        paste("BIDp",i,sep=""),paste("BIDv",i,sep="")) }
colnames(data) <- colnms
```

```
# clean up
rm('dataOB', 'dataM')
```

4.4.4 Application: Visualization of LOB and Order Flow

The purpose of the LOB and order flow visualization is to help researchers intuitively understand the basic principles of order-driven trading by showing a sequence of changes in the LOB associated with the incoming order events. The upper left box in Figure 4.7 contains basic information about the displayed ticker and current time, as well as a comment field, which interprets the situation at the given point in time, thus helping to understand the principles better. The box labeled "Order Flow" shows several orders around the current order (violet color), as recorded in the corresponding output file. The box Limit Order Book (Table) box contains rows from the LOB that correspond to the orders displayed in the Order Flow box. The Limit Order Book (Table) box contains prices and depth (number of shares) at three levels of bid and ask. The box with the title "Limit Order Book (Graph)" contains the same information (for one row) visualized; depth (number of shares) at the three levels of displayed order book. Note that this graph contains only the current state of the LOB corresponding to the violet row in the previous two boxes. Finally, the large graph on the right displays the incoming orders and the continuity of BID and ASK price levels against the time axis. Note that this graph does not contain information about the size of incoming orders nor the cumulated depth. But in combination, both graphs contain all information from the tables and thus provide current snapshots, as well as the dynamics of the LOB in the sample period.

4.5 Order Tracer

We also implement a module called `Order Tracer` to extract information, i.e. cancellation, execution and deletion, for single limit orders. The procedure is similar to the `LOB Constructor`.

4.5.1 Implementation of Order Tracer

Reusing the classes designed for the `LOB Constructor`, we implement another application by adding one additional controlling class - the `TraceRecorder`. It passes the message data and completes information from the underlying limit orders exactly in the same way as the `LOB Constructor`. However, rather than using the information to update the LOB, it simply records the types of events and limit orders temporarily in a block of memory called `tracePool`. The constructed data are then saved by instances of `LimitOrderWriter` and `TraceWriter`. Figure 4.8 contains our final class diagram.

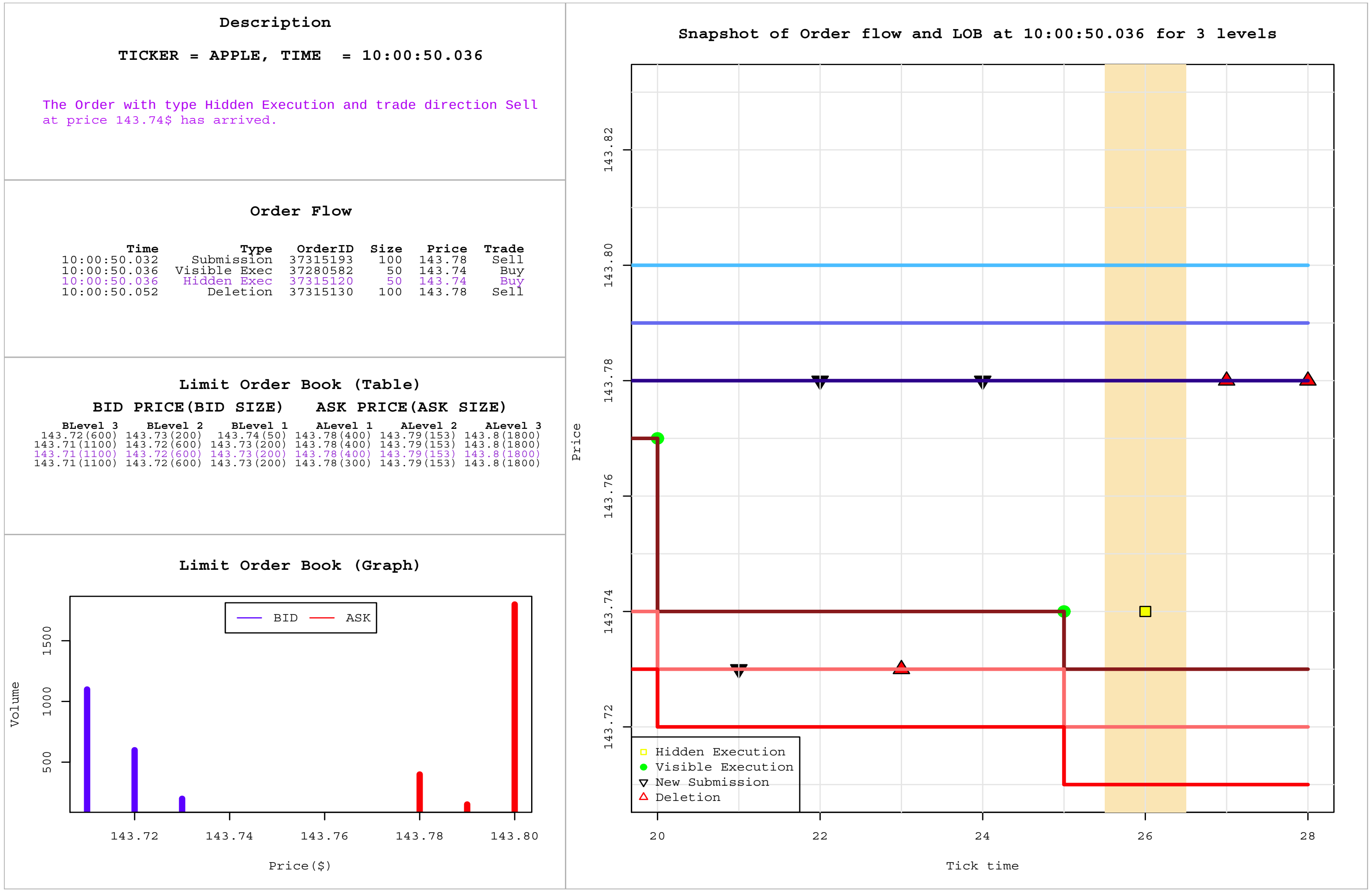


Figure 4.7: The visualization of LOB and order flow. The time X-axis in the right graph does not correspond to the real time, but rather to "tick time", i.e. events in this graph are equidistant with respect to the X-axis regardless of their actual distance in time. The Y-axis "Price" is scaled realistically to represent the actual quotes in the LOB. The current order event and LOB instance are highlighted by violet color in the left graph and by orange color in the right graph.

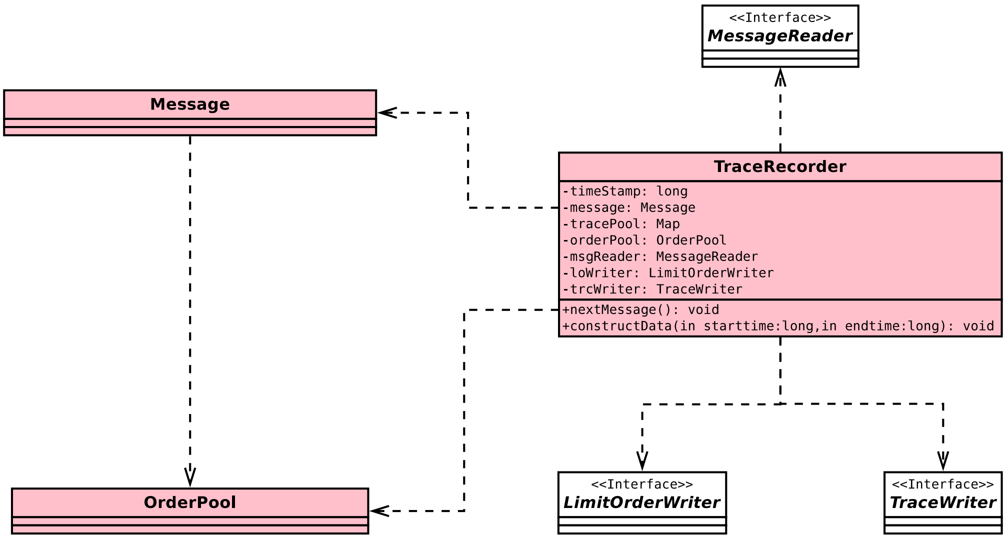


Figure 4.8: Class diagram of trace construction. A `TraceRecorder` object contains unique `OrderPool` objects, which are updated by incoming messages corresponding to the `Message` objects. Also, it must include at least one `MessageReader` object for reading the input, a `LimitOrderWriter` object and a `TraceWriter` objects for writing the output into the file system.

4.5.2 Output of Order Tracer

Similar to the `LOB Constructor`, the `Order Tracer` generates two files with the same number of rows. The first output file classifies events as cancellations or executions and contains the following six columns (see a sample in Table 4.4).

- `submission time`: the time when the corresponding limit order was submitted; milliseconds after midnight.
- `time`: time stamp of the event; milliseconds after midnight.
- `size`: change of limit order size; number of shares.
- `execution`: a dummy variable indicating whether the event corresponds to a limit order execution (1) or cancellation (0).
- `earlier exe.`: a dummy variable indicating whether the event corresponds to a limit order that has been partially executed earlier (1) or not (0).

The second output file contains information about the corresponding limit orders. Table 4.5 contains a segment of this file. There are seven variables:

- `order ID`: a unique number generated by the exchange identifying the limit order.

Table 4.4
Order event data generated by the Order Tracer.

The sample contains a segment of data for GOOG from October 1st 2010. The time variable is in milliseconds after midnight. Size is the effective size, i.e. cancelled (executed) quantity of cancellation (execution) events. Variable Execution indicates the type of event (execution or cancellation). Variable Earlier exe. indicates the first execution by zero. The underlying limit order information is shown in Table 4.5.

Submission Time	Time	Size	Execution	Earlier exe.
36502739	36502742	100	0	0
36502743	36502743	100	0	0
36502554	36502744	100	0	0
36502744	36502745	100	0	0
36502726	36502745	100	0	0
36502734	36502751	100	0	0
36502745	36502755	100	1	0
36502742	36502755	123	1	0
36502755	36502755	100	1	0
36502745	36502756	100	0	0
36502730	36502756	100	0	0
36502742	36502757	177	1	1
36502750	36502757	100	1	0
36502741	36502757	23	1	0
36502741	36502757	77	1	1
36502737	36502757	100	1	0
36502047	36502757	100	0	0
36502761	36502761	100	1	0
36502757	36502761	300	0	0
36502761	36502761	100	0	0

- `order size`: original order size at submission.
- `remaining size`: remaining size of the order after an order event.
- `price`: price of the limit order.
- `trade direction`: 1 for buy limit orders, -1 for sell limit orders.
- `hidden`: a dummy variable indicating whether the order is hidden (1) or not (0).
- `MPID`: Market Participant ID of the trader who submitted the order, null if unknown.

Since we organize the constructed data in such a way that the two output files contain identical number of observations, we can easily load and merge them using statistical software, such as Matlab and R. Here is an example of Matlab code.

Table 4.5
Limit order data generated by Order Tracer.

The sample contains a segment of data for GOOG from October 1st 2010. Variable Rem. Size is the remaining share quantity of the limit order after an event. The Price variable is in 0.01 cent. Some limit orders may contain additional MPID information identifying the submitters. Events recorded for individual limit orders are shown on Table 4.4.

Order ID	Order Size	Rem. Size	Price	Trade Dir.	Hidden	MPID
57916505	100	0	5273000	1	0	null
57916618	100	0	5273100	1	0	null
57913456	100	0	5261200	1	0	HDSN
57916642	100	0	5273300	1	0	null
57916188	100	0	5271700	1	0	null
57916358	100	0	5271700	1	0	null
57916698	100	0	5273400	1	0	null
57916588	300	177	5273300	1	0	null
57916928	100	0	5273400	1	0	null
57916679	100	0	5273000	1	0	null
57916279	100	0	5221600	1	0	NMRA
57916588	300	0	5273300	1	0	null
57916814	100	0	5273300	1	0	null
57916545	100	77	5273100	1	0	null
57916545	100	0	5273100	1	0	null
57916448	100	0	5273000	1	0	null
57899910	100	0	5275100	-1	0	null
57917068	100	0	5273300	1	0	null
57917003	300	0	5265700	1	0	null
57917079	100	0	5275000	-1	0	null

An Example of Matlab Code Loading Order Trace Data

```
% load event data
trace_data=load('GOOG_20101001_trace.csv']);

% load order data
fid=fopen(['GOOG_20101001_order.csv']);
orders=textscan(fid,'%f,%f,%f,%f,%f,%f,%s');
fclose(fid);
order_data=[orders{1} orders{2} orders{3} orders{4} orders{5} orders{6}];

% merge the datasets
mydata=[traceData orderData];

% clean up
```

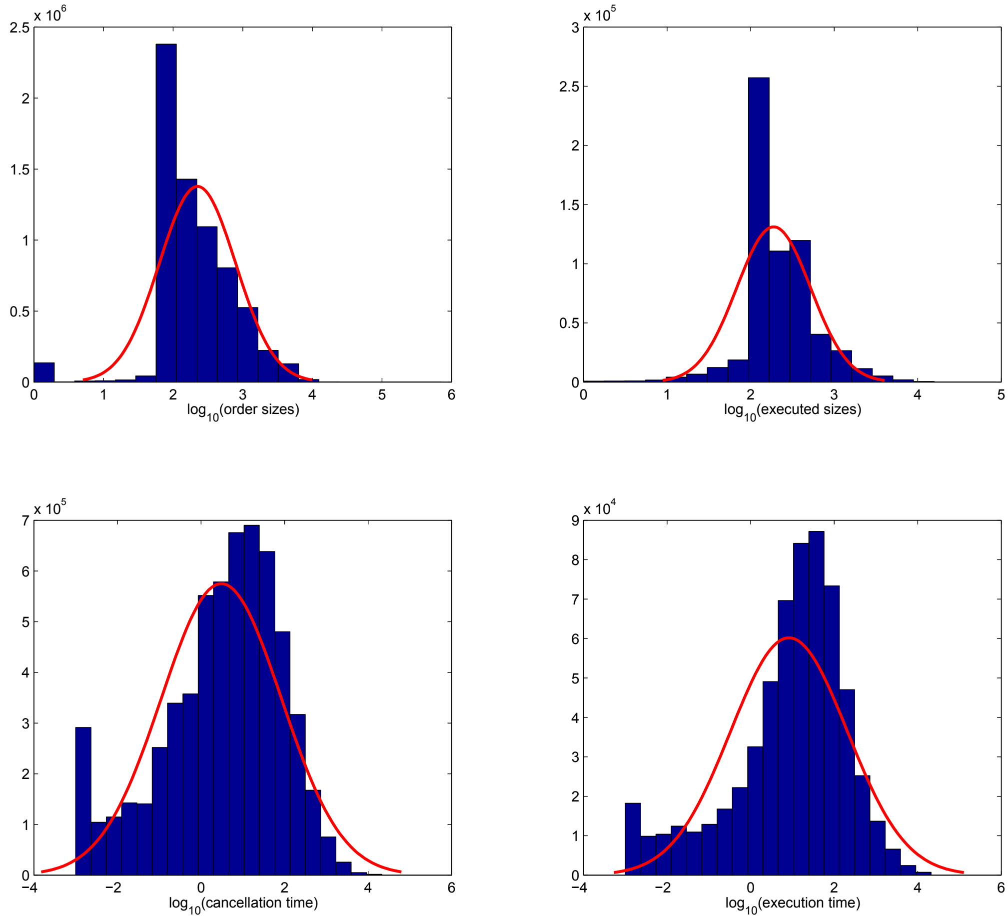


Figure 4.9: Histogram of size, execution quantity, cancellation time and execution time for limit orders. The red line represents kernel density estimates. Zero cancellation time and execution are discarded. Trading data for Microsoft Corp. on NASDAQ in October, 2010

```
clear trace_data order_data orders;
```

4.5.3 Application: Main Characteristics of Limit Orders

Figure 4.9 shows a simple analysis of the characteristics of limit orders using the output of the `Order Tracer`. Calculation of the order sizes, execution quantities and life time of limit orders are trivial tasks when using the data illustrated in Table 4.4 and 4.5. Nevertheless, we can make a few interesting empirical observations: 1) Market participants submit a huge number of limit orders with small sizes. Indeed, we find that most limit orders are of size 100, which is the size of a round lot in NASDAQ. 2) Only a small proportion of limit orders are executed and the execution quantity is small. 3) Most of limit orders are cancelled shortly after the submission. 4) Execution time is typically longer than cancellation time. More detailed analysis of order flow properties and limit order characteristics can be found in Hautsch and Huang (2012a).

4.6 Conclusion

System LOBSTER is designed to meet the three basic requirements for constructing datasets for empirical studies on limit order markets. First, it is sufficiently efficient and *fast*, requiring only seconds or minutes to fulfill standard requests. The system is web-based and very intuitive with a *user-friendly* interface allowing researchers to fully focus on research rather than spend time preparing data. Third, the system was programmed using object-oriented programming language and intentionally designed to allow for easy extensions, which makes it very *versatile*. New modules of the LOB Constructor and Order Tracer have been already added to the system.

Of course, any meaningful extension must be based on a sound research idea. To facilitate communication with other researchers, we created a forum (`http://lobster.wiwi.hu-berlin.de/forum/`) focused on modelling order book and order flow. It is a slowly but steadily growing source of information, not only about LOBSTER, but about everything related to order-driven markets. We expect that it will become a rich and comprehensive pool of references for academic researchers, which will help them to accelerate the initial stages of their research projects and help us further develop the system. The feedback we are receiving from students and researchers pioneering with our system has already proven to be essential in the development of the system.

Bibliography

Aitken, M., H. Berkman, and D. Mak, 2001, The use of undisclosed limit orders on the Australian Stock Exchange, *Journal of Banking and Finance* 25, 1589 – 1603.

Anand, A., and D. Weaver, 2004, Can order exposure be mandated?, *Journal of Financial Markets* 7, 405 – 426.

Angel, J., L. Harris, and C. Spatt, 2010, Equity trading in the 21st century, Working Paper FBE 09-10 Marshall School of Business.

Bessembinder, H., M. Panayides, and K. Venkataraman, 2009, Hidden liquidity: an analysis of order exposure strategies in electronic stock markets, *Journal of Financial Economics* 94, 361 – 383.

Biais, B., P. Hillion, and C. Spatt, 1995, An empirical analysis of the limit order book and the order flow in intermediation, *Journal of Finance* 50, 1655 – 1689.

Bloomfield, R., M. O'Hara, and G. Saar, 2005, The "make or take" decision in an electronic market: Evidence on the evolution of liquidity, *Journal of Financial Economics* 75, 165 – 199.

Boulatov, A., and T. George, 2008, Securities trading when liquidity providers are informed, Working paper University of Houston.

Buti, S., and B. Rindi, 2011, Undisclosed orders and optimal submission strategies in a dynamic limit order market, EFA 2008 Athens Meeting Paper, AFA 2009 San Francisco Meeting Paper.

Cebiroglu, G., and U. Horst, 2011, Optimal display of iceberg orders, SFB 649 Discussion Paper 2011-057 Humboldt Universität zu Berlin.

Chiarella, C., G. Iori, and J. Perelló, 2009, The impact of heterogeneous trading rules on the limit order book and order flows, *Journal of Economic Dynamics & Control* 33, 525 – 537.

Cont, R., A. Kukanov, and S. Stoikov, 2011, The price impact of order book events, *SSRN eLibrary.*

De Winne, R., and C. D'Hondt, 2007, Hide-and-seek in the market: placing and detecting hidden orders, *Review of Finance* 11, 663 – 692.

Dufour, A., and R. Engle, 2000, Time and the impact of a trade, *Journal of Finance* 55, 2467 – 2498.

DuMouchel, W., 1994, Hierarchical Bayes linear models for meta-analysis, Technical Report 27 National Institute of Statistical Sciences.

Easley, D., N. M. Kiefer, and M. O'Hara, 1997, One day in the life of a very common stock, *Review of Financial Studies* 10, 805 – 835.

Easley, D., and M. O'Hara, 1992, Time and the process of security price adjustment, *Journal of Finance* 47, 577 – 605.

Eisler, Z., J. Bouchaud, and J. Kockelkoren, 2011, The price impact of order book events: Market orders, limit orders and cancellations, *Quantitative Finance* Forthcoming.

Engle, R., and A. Patton, 2004, Impact of trades in an error-correction model of quote prices, *Journal of Financial Markets* 7, 1 – 25.

Eom, K., E. Lee, and K. Park, 2009, Microstructure-based manipulation: Strategic behavior and performance of spoofing traders, *SSRN eLibrary.*

Escribano, A., and R. Pascual, 2006, Asymmetries in bid and ask responses to inonovation in the trading process, *Empirical Economics* 30, 913 – 946.

Esser, A., and B. Mönch, 2007, The navigation of an iceberg: The optimal use of hidden orders, *Finance Research Letters* 4, 68 – 81.

Fleming, M., and B. Mizrach, 2009, The microstructure of a U.S. treasury ecn: The BrockerTec platform, Staff Report 381 Federal Reserve Bank of New York.

Foucault, T., 1999, Order flow composition and trading costs in a dynamic limit order market, *Journal of Financial Markets* 2, 99 – 124.

Frey, S., and P. Sandås, 2009, The impact of iceberg orders in limit order books, CFR Working Paper 09-06 University of Cologne, Centre for Financial Research (CFR).

Glosten, L., and P. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71 – 100.

Gredenhoff, M., and T. Jacobson, 2001, Bootstrap testing linear restrictions on cointegrating vectors, *Journal of Business and Economic Statistics* 19, 63 – 72.

Griffiths, M., B. Smith, D. Turnbull, and R. White, 2000, The costs and determinants of order aggressiveness, *Journal of Financial Economics* 56, 65 – 88.

Hall, T., and N. Hautsch, 2005, Order aggressiveness and order book dynamics, *Empirical Economics* 30, 973 – 1005.

Hansen, P., and A. Lunde, 2006, Realized variance and market microstructure noise, *Journal of Business and Economic Statistics* 24, 127 – 161.

Härdle, W., N. Hautsch, and A. Mihoci, 2009, Modelling and forecasting liquidity supply using semiparametric factor dynamics, Discussion Paper 2009-018 Collaborative Research Center 649 "Economic Risk", Humboldt-Universität zu Berlin.

Harris, L., 1996, Does a large minimum price variation encourage order exposure, Working paper Marshall School of Business, University of Southern California.

——— , 1997, Order exposure and parasitic traders, Working paper Marshall School of Business, University of Southern California.

Hasbrouck, J., 1991, Measuring the information content of stock trades, *Journal of Finance* 46, 179 – 207.

——— , and G. Saar, 2009, Technology and liquidity provision: The blurring of traditional definitions, *Journal of Financial Markets* 12, 143 – 172.

Haug, A., 2002, Testing linear restrictions on cointegrating vectors: sizes and powers of Wald and likelihood ratio tests in finite samples, *Econometric Theory* 18, 505 – 524.

Hautsch, N., and R. Huang, 2012a, Limit order flow, market impact, and optimal order sizes: Evidence from NASDAQ TotalView-ITCH data, in F. Abergel, J.-P. Bouchaud, T. Foucault, C. Lehal, and M. Rosenbaum, ed.: *Market Microstructure: Confronting Many Viewpoints – Conference Proceedings* pp. 137 – 161. John Wiley & Sons, Ltd.

——— , 2012b, The market impact of a limit order, *Journal of Economic Dynamics & Control* 36, 501 – 5022.

——— , 2012c, On the dark side of the market: Identifying and analyzing hidden order placement, Discussion Paper 2012-14, CRC 649 Humboldt Universität zu Berlin, Germany.

Hendershortt, T., C. Jones, and A. Menkveld, 2010, Does algorithmic trading improve liquidity, Working paper Haas School of Business, University of California Berkeley.

Huang, R., and T. Polak, 2011, LOBSTER: The limit order book reconstructor, Discussion paper School of Business and Economics, Humboldt Universität zu Berlin `http://lobster.wiwi.hu-berlin.de/Lobster/LobsterReport.pdf`.

Johansen, S., 1991, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica* 59, 1551 – 80.

——— , 1995, *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models* (Oxford University Press).

——— , and K. Juselius, 1990, Maximum likelihood estimation and inference on cointegration - with applications to the demand for money, *Oxford Bulletin of Economics and Statistics* 52, 169 – 210.

Johansen, S., and M. Nielsen, 2010, Likelihood inference for a fractionally cointegrated vector autoregressive model, Research Paper 2010-24 CREATES.

Jorda, O., 2005, Estimation and inference of impulse response by local projections, *American Economic Review* 95, 161 – 182.

Kwiatkowski, D., P. Phillips, P. Schmidt, and Y. Shin, 1992, Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics* 54, 159 – 178.

Lee, C., and M. Ready, 1991, Inferring trade direction from intraday data, *Jounal of Finance* 46, 733 – 746.

Ljung, G., and G. Box, 1978, On a measure of lack of fit in time series models, *Biometrika* 65, 297 – 303.

Lo, I., and S. Sapp, 2006, A structural error-correction model of best prices and depths in the foreign exchange limit order market, Working Paper 06-8 Bank of Canada.

Lütkepohl, H., 1990, Asymptotic distributions of impulse response functions and forecast error variance decompositions of vector autoregressive models, *Review of Economics and Statistics* 72, 116 – 125.

——— , and H. Reimers, 1992, Impulse response analysis of cointegrated systems, *Journal of Economic Dynamics & Control* 16, 53 – 78.

McKelvey, R., and W. Zaviona, 1975, A statistical model for the analysis of ordinal level dependent variables, *Journal of Mathematical Sociology* 4, 103–120.

Mike, S., and J. Farmer, 2008, An empirical behavioral model of liquidity and volatility, *Journal of Economic Dynamics & Control* 32, 200 – 234.

Moinas, S., 2010, Hidden limit orders and liquidity in order driven markets, working paper 10-147 Toulouse School of Economics.

NASDAQ, 2008, NASDAQ stock market rules, http://nasdaq.cchwallstreet.com/.

Pardo Tornero, A., and R. Pascual, 2007, On the hidden side of liqudity, Available at SSRN: http://ssrn.com/abstract=459000 or doi:10.2139/ssrn.459000.

Parlour, C., 1998, Price dynamics and limit order markets, *Review of Finance Studies* 11, 789 – 816.

———, and D. Seppi, 2008, Limit order markets: A survey, in A. Thakor, and A. Boot, ed.: *Handbook of Financial Intermediation & Banking* pp. 63 – 96.

Potters, M., and J. Bouchaud, 2003, More statistical properties of order books and price impact, *Physica A* 324, 133 – 140.

Rosu, I., 2010, Liquidity and information in order driven markets, Working paper HEC Paris Available at SSRN: http://ssrn.com/abstract=1286193.

Russell, J., and T. Kim, 2010, A new model for limit order book dynamics, in T. Bollerslev, J. Russell, and M. Watson, ed.: *Volatility and Time Series Econometrics, Essays in Honor of Robert Engle* (Oxford Univerty Press).

Said, E., and D. Dickey, 1984, Testing for unit roots in autoregressive moving average models of unknown order, *Biometrika* 71, 599 – 607.

Securities and Exchange Commission, 2010, Concept release on equity market structure, *Federal Register* 75, 3594 – 3614.

Tuttle, L., 2006, Hidden orders, trading costs, and information, Unpublished working paper University of Kansas, Lawrence, KS.

White, H., 1980, A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, *Econometrica* 48, 817 – 838.

Yamamoto, R., 2011, Order aggressiveness, pre-trade transparency, and long memory in an order-drivern market, *Journal of Economic Dynamics & Control* 35, 1938 – 1963.

Appendix A

A.1 Adaptive time window matching algorithm

In our database, trade data and order book data are recorded in separate files stemming from different recording systems. As a result, the time stamps in the two data sets have different time distances to exchange time. In accordance with the institutional settings of Euronext, we design an adaptive time window matching algorithm which contains three main steps.

Step 1 Exact matching. The algorithm picks up a time stamp of a trade and opens a specified time window, e.g. $[-10, 10]$ seconds around this time stamp. Then, a procedure picks every order book record in this time window and performs the following analysis: If (i) the trade price equals to the best bid (ask) price and the difference of the best bid (ask) size between this order book record and the previous one equals to the trade size or (ii) the trade price equals to the previous best bid (ask) price, the best bid (ask) size equals to the trade size and the best bid (ask) price decreases (increases), it matches this order book record with the corresponding trade and records the delay time between the trade and the order book. If no match is achieved for all order book records in the time window, the trade remains to be unmatched.

Step 2 Inexact matching. The algorithm picks up an unmatched trade record's time stamp and opens a time window of size which is twice the average delay time computed in Step 1. If (i) the trade price equals the best bid (ask) price and the best bid (ask) size is less than the previous one or (ii) the best bid (ask) price decreases (increases), it matches the trade with the current order book. If no match is achieved for all order book records in the time window, the trade remains to be unmatched.

Step 3 Round time matching. The algorithm picks up an unmatched trade and matches it with an order book record that is closed to the trade's time stamp plus the average delay time.

Figure A.1 gives the histogram of the delay time between trades and their corresponding order book records. The delay time is computed in Step 1 by which 1609 (sub-)trades

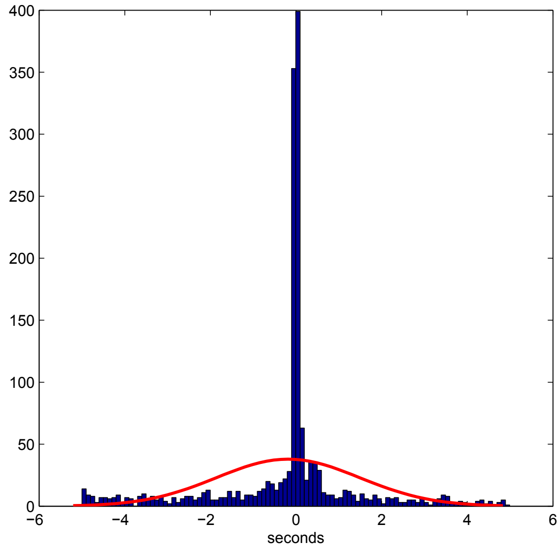


Figure A.1: Histogram of the delay time between the trade and its corresponding order book record. Trading of Fortis, Euronext Amsterdam on August 1st, 2008.

have been exactly matched with their corresponding order books inside a $[-5, 5]$ second time windows. The average delay time is -0.185 seconds, i.e., trades are on average recorded 185 milliseconds before the corresponding order book.

A.2 FIML estimator for cointegrating vectors

Model (1.1) is estimated by the Full Information Maximum Likelihood (FIML) estimator proposed by Johansen (1991) and Johansen and Juselius (1990). Let $z_{0t} := \Delta y_t$, and $z_{1t} := y_{t-1}$. Further let z_{2t} be the vector of stacked variables,

$$z_{2t} := (\Delta y_{t-1}, \cdots, \Delta y_{t-p+1}, x_{t-1}, \cdots, x_{t-s}, 1)'$$

with corresponding parameter vector $\Gamma = (\Gamma_1, \ldots, \Gamma_{p-1}, \mu)$. Define the product moment matrices

$$M_{ij} := T^{-1} \sum_{t=1}^{T} z_{it} z'_{jt}, \quad i, j = 0, 1, 2,$$

where T is the number of observations. Moreover, let

$$S_{ij} := M_{ij} - M_{i2} M_{22}^{-1} M_{2j}.$$

We then solve the generalized eigenvalue problem

$$\left|\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}\right| = 0$$

for the eigenvalues $1 > \hat{\lambda}_1 > \cdots > \hat{\lambda}_K > 0$ and corresponding eigenvector $\hat{V} = (\hat{v}_1, \cdots, \hat{v}_K)$ which is normalized by $\hat{V}S_{11}\hat{V} = I_K$. Johansen's (1991) trace test or maximum eigenvalue test can be used to determine the underlying cointegration rank r. Under the hypothesis that there exist r cointegration relations, the $K \times r$ cointegrating matrix β is estimated by

$$\hat{\beta} = (\hat{v}_1, \ldots, \hat{v}_r)$$

with corresponding maximized log-likelihood function

$$l_{\max}(\hat{\beta}) = -\frac{T}{2}\left(\ln|S_{00}| + \sum_{i=1}^{r}\ln(1-\hat{\lambda}_i)\right). \tag{A.2.1}$$

The magnitude of $\hat{\lambda}_i$ can be interpreted as a measure of the stationarity of the product $\hat{\beta}_i' y_t$. The larger $\hat{\lambda}_i$, the closer the stochastic properties of the underlying relationship to that of a stationary process. The parameters α and Γ are estimated by OLS after inserting $\hat{\beta}$ into equation (1.1) and computing $\hat{\Sigma}_u$ as $\hat{\Sigma}_u = S_{00} - \hat{\alpha}\hat{\alpha}'$.

When some of the cointegration relations are known, we can partition the cointegrating matrix as

$$\beta = (b, \varphi)$$

where b contains known cointegrating vectors and φ contains the unknown ones. Denote

$$S_{ij.b} = S_{ij} - S_{i1}(b'S_{11}b)^{-1}S_{1j}.$$

Then, φ can be estimated similarly by solving the generalized eigenvalue problem

$$\left|\lambda b_\perp' S_{11.b} b_\perp - b_\perp' S_{10.b} S_{00.b}^{-1} S_{01.b} b_\perp\right| = 0$$

A.3 Transform estimates from VECM to VAR

Following Lütkepohl and Reimers (1992), the parameters of equation (1.1) can be easily transformed to equation (1.2). In this context, we define a transformation matrix

$$D = \underbrace{\left[\begin{array}{ccccccc|c}
I_K & 0 & 0 & & \cdots & 0 & 0 & 0 \\
I_K & -I_K & 0 & & \cdots & 0 & 0 & 0 \\
0 & I_K & -I_K & \ddots & & 0 & 0 & 0 \\
\vdots & \vdots & & \ddots & \ddots & & \vdots & \vdots \\
0 & 0 & 0 & & \ddots & 0 & 0 & 0 \\
0 & 0 & 0 & & \cdots & -I_K & 0 & 0 \\
0 & 0 & 0 & & \cdots & I_K & -I_K & 0 \\
\hline
 & & & 0 & & & & 1
\end{array}\right]}_{(KP+1)\times(KP+1)}$$

such that

$$[A_1, \cdots, A_p, \mu] = [\alpha\beta', \Gamma]D + J^*, \tag{A.3.2}$$

where $J^* := [I_K : 0 : \cdots : 0]$ is a $K \times (Kp + 1)$ matrix. The theorem below provides a consistent estimator of A and B:

Theorem 1 (Lütkepohl and Reimers, 1992)**.** *Let $\hat{\alpha}$, $\hat{\beta}$, $\hat{\Gamma}$ and $\hat{\Sigma}_u$ denote the FIML estimates of the parameters of model* (1.1)*. Moreover, $\hat{A}_1, \cdots, \hat{A}_p, \hat{\mu}$ are computed by the transformation in* (A.3.2)*. Then,*

$$\sqrt{T}\left[\mathrm{vec}(\hat{A}_1, \cdots, \hat{A}_p, \hat{\mu}) - \mathrm{vec}(A_1, \cdots, A_p, \mu)\right] \xrightarrow{d} \mathcal{N}(0, \Sigma_{AB}), \tag{A.3.3}$$

where

$$\Sigma_{AB} = D' \begin{bmatrix} \beta & 0 \\ 0 & I_{K(p-1)+1} \end{bmatrix} \Omega^{-1} \begin{bmatrix} \beta' & 0 \\ 0 & I_{K(p-1)+1} \end{bmatrix} D \otimes \Sigma_u,$$

$$\Omega = \mathrm{plim}\, \frac{1}{T} \begin{bmatrix} \beta' M_{11} \beta & \beta' M_{12} \\ M_{21}\beta & M_{22} \end{bmatrix}$$

are consistently estimated by

$$\hat{\Sigma}_{AB} = D' \begin{bmatrix} \hat{\beta} & 0 \\ 0 & I \end{bmatrix} \hat{\Omega}^{-1} \begin{bmatrix} \hat{\beta}' & 0 \\ 0 & I \end{bmatrix} D \otimes \hat{\Sigma}_u$$

$$\hat{\Omega} = \begin{bmatrix} \hat{\beta}' M_{11} \hat{\beta} & \hat{\beta}' M_{12} \\ M_{21}\hat{\beta} & M_{22} \end{bmatrix}.$$

Appendix B

B.1 Asymptotic Distribution of Marginal Effects

The asymptotic covariance of marginal dummy effects is straightforwardly computed using the delta method. Let $\boldsymbol{\theta} = [\boldsymbol{\beta}', \gamma_1, \ldots, \gamma_{J-1}]'$ be the vector of $(K+J-1)$ unknown parameters, $\widehat{\boldsymbol{\theta}}$ denotes the maximum likelihood estimator with $V \equiv \text{Asy.Var}[\widehat{\boldsymbol{\theta}}]$ being its $(K+J-1) \times (K+J-1)$ asymptotic covariance matrix. Then, the asymptotic covariance matrix of the corresponding marginal effects is given by

$$\text{Asy.Var}[\widehat{\boldsymbol{q}}_j] = \left[\frac{\partial \widehat{\boldsymbol{q}}_j}{\partial \widehat{\boldsymbol{\theta}}'}\right] V \left[\frac{\partial \widehat{\boldsymbol{q}}_j}{\partial \widehat{\boldsymbol{\theta}}'}\right]', \tag{B.1.1}$$

where the derivatives of $\widehat{\boldsymbol{q}}_j$ with respect to $\widehat{\boldsymbol{\theta}}'$ are

$$\begin{aligned}
\frac{\partial \widehat{\boldsymbol{q}}_1}{\partial \widehat{\boldsymbol{\theta}}'} &= \left[\phi_1(I_K - z_1\widehat{\boldsymbol{\beta}}\boldsymbol{x}'),\ \phi_1 z_1\widehat{\boldsymbol{\beta}},\ \mathbf{0},\ \cdots,\ \mathbf{0}\right], \\
\frac{\partial \widehat{\boldsymbol{q}}_2}{\partial \widehat{\boldsymbol{\theta}}'} &= \left[(\phi_1 - \phi_2)I_K + (\phi_1 z_1 - \phi_2 z_2)\widehat{\boldsymbol{\beta}}\boldsymbol{x}',\ -\phi_1 z_1\widehat{\boldsymbol{\beta}},\ \phi_2 z_2\widehat{\boldsymbol{\beta}},\ \cdots,\ \mathbf{0}\right], \\
&\vdots \\
\frac{\partial \widehat{\boldsymbol{q}}_J}{\partial \widehat{\boldsymbol{\theta}}'} &= \left[\phi_{J-1}(I_K + z_{J-1}\widehat{\boldsymbol{\beta}}\boldsymbol{x}'),\ \mathbf{0},\ \mathbf{0},\ \cdots,\ -\phi_{J-1} z_{J-1}\widehat{\boldsymbol{\beta}}\right],
\end{aligned} \tag{B.1.2}$$

with I_K denoting a $K \times K$ identity matrix, $\mathbf{0}$ is a $(K \times 1)$ zero vector and z_j and ϕ_j given by $z_j \equiv \gamma_j - \widehat{\boldsymbol{\beta}}'\boldsymbol{x}$ and $\phi_j \equiv \phi(z_j)$. Then,

$$\text{Asy.Var}[\Delta\widehat{F}_j] = \left[\frac{\partial \Delta\widehat{F}_j}{\partial \widehat{\boldsymbol{\theta}}'}\right] V \left[\frac{\partial \Delta\widehat{F}_j}{\partial \widehat{\boldsymbol{\theta}}'}\right]', \tag{B.1.3}$$

where

$$\frac{\partial \Delta\widehat{F}_j}{\partial \widehat{\boldsymbol{\theta}}'} = \left.\frac{\partial \widehat{F}_j}{\partial \widehat{\boldsymbol{\theta}}'}\right|_{\boldsymbol{x}_{(x_d=1)}} - \left.\frac{\partial \widehat{F}_j}{\partial \widehat{\boldsymbol{\theta}}'}\right|_{\boldsymbol{x}_{(x_d=0)}}$$

with

$$
\begin{aligned}
\frac{\partial \widehat{F}_1}{\partial \widehat{\boldsymbol{\theta}}'} &= \left[\phi_1 \boldsymbol{x}',\ \phi_1,\ \mathbf{0},\ \cdots,\ \mathbf{0}\right], \\
\frac{\partial \widehat{F}_2}{\partial \widehat{\boldsymbol{\theta}}'} &= \left[(\phi_2 - \phi_1) \boldsymbol{x}',\ -\phi_1,\ \phi_2,\ \cdots,\ \mathbf{0}\right], \\
&\vdots \\
\frac{\partial \widehat{F}_J}{\partial \widehat{\boldsymbol{\theta}}'} &= \left[\phi_{J-1} \boldsymbol{x}',\ \mathbf{0},\ \mathbf{0},\ \cdots,\ -\phi_{J-1}\right].
\end{aligned}
$$

B.2 Histogram of Significant Ordered Probit Estimates

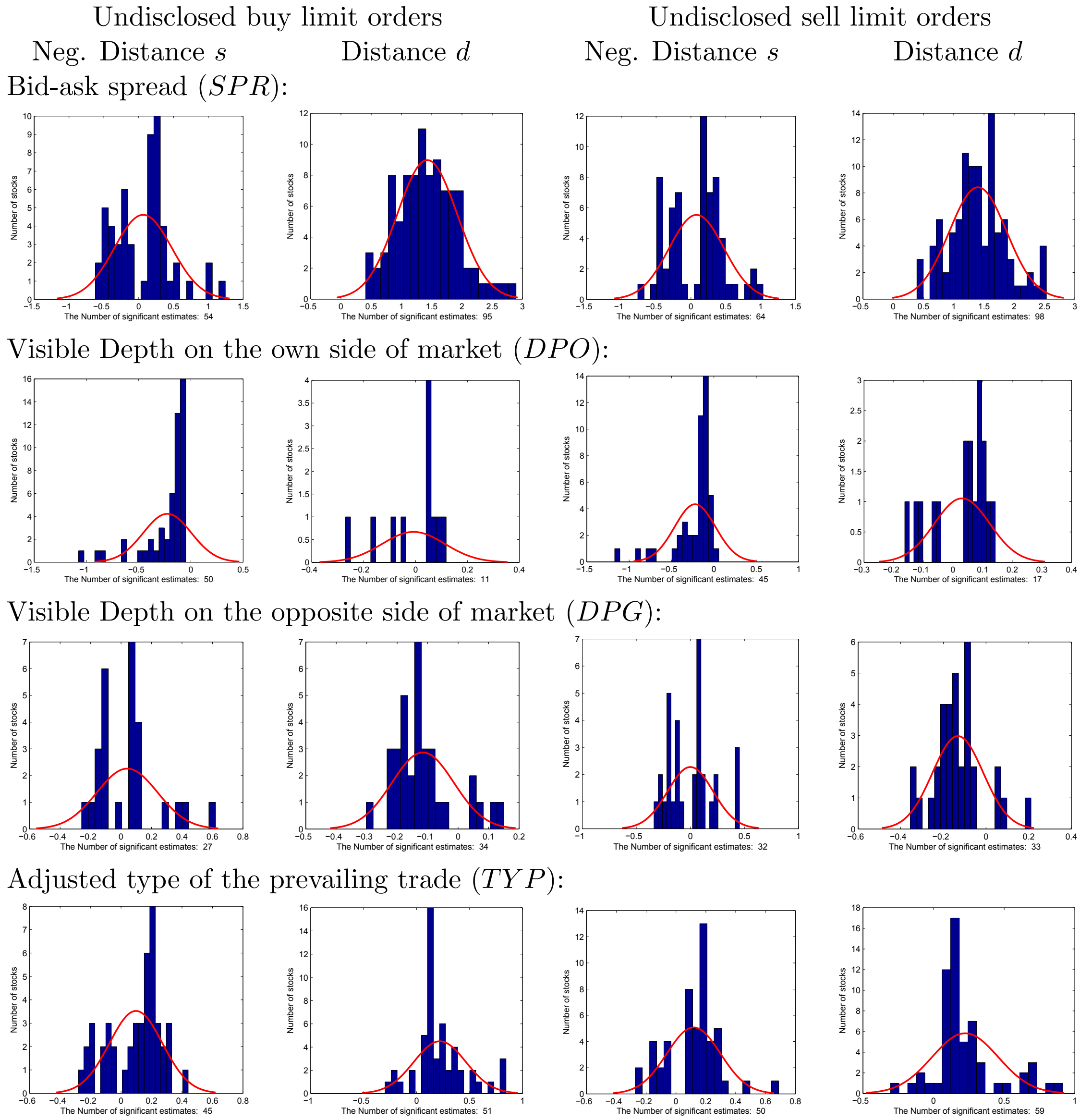


Figure B.1: The histogram of significant estimates. Ordered probit estimates of hidden order locations on the bid and ask side depending on categorized measures s and d as discussed in Section 3.3.4. The order aggressiveness is declining with the category label, thus negative coefficients are associated with increasing aggressiveness. Based on 99 NASDAQ stocks during the period of October 2010 with 21 trading days, the histogram shows estimates significantly different from zero at 5% α-level. TYP is adjusted such that it equals 1 when the prevailing trade consumes the own-side liquidity, -1 otherwise.

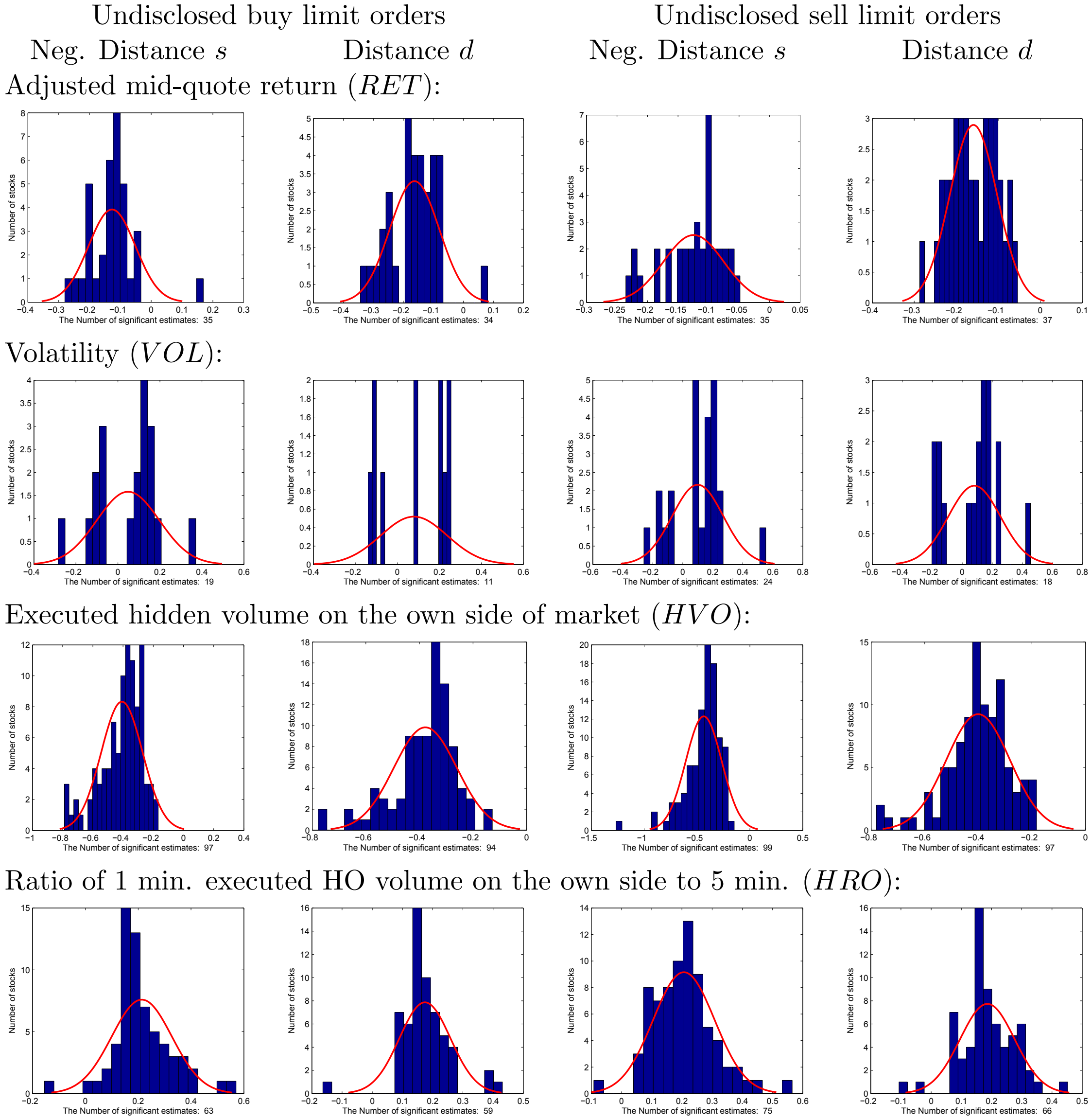
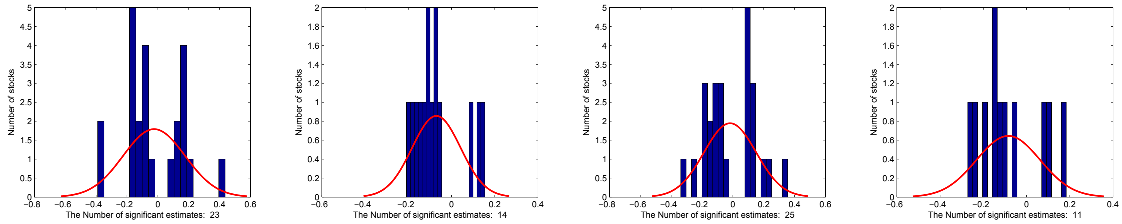


Figure B.2: The histogram of significant estimates. Ordered probit estimates of hidden order locations on the bid and ask side depending on categorized measures s and d as discussed in Section 3.3.4. The order aggressiveness is declining with the category label, thus negative coefficients are associated with increasing aggressiveness. Based on 99 NASDAQ stocks during the period of October 2010 with 21 trading days, the histogram shows estimates significantly different from zero at 5% α-level. RET is adjusted such that it equals to positive value when mid-quotes move away the own side market.

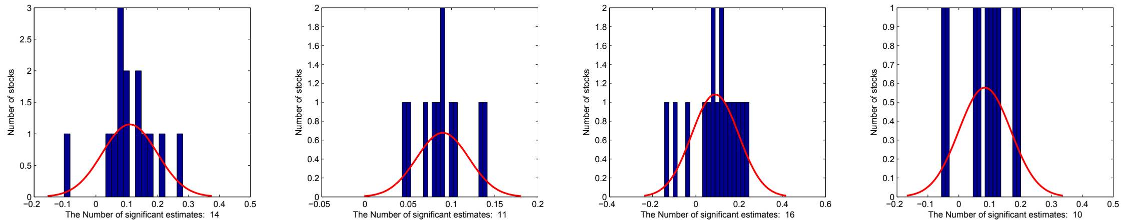
Undisclosed buy limit orders | Undisclosed sell limit orders

Neg. Distance s | Distance d | Neg. Distance s | Distance d

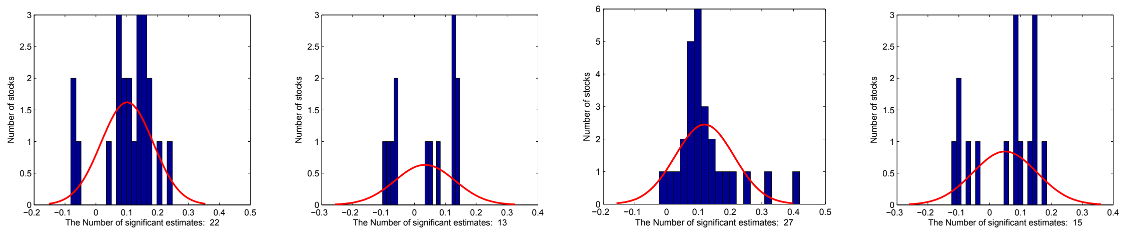
Executed hidden volume on the opposite side of market (HVG)



Ratio of 1 min. executed HO volume on the opposite side to 5 min. (HRG):



Number of "low frequency" limit order updating the own-side quote (ALO)



Number of "low frequency" limit order updating the opposite-side quote (ALG):

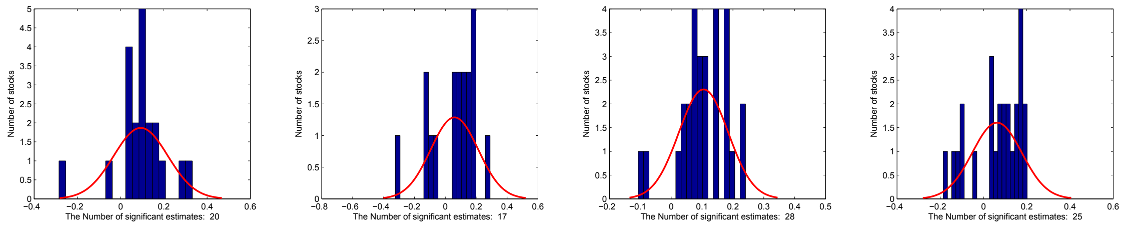


Figure B.3: The histogram of significant estimates. Ordered probit estimates of hidden order locations on the bid and ask side depending on categorized measures s and d as discussed in Section 3.3.4. The order aggressiveness is declining with the category label, thus negative coefficients are associated with increasing aggressiveness. Based on 99 NASDAQ stocks during the period of October 2010 with 21 trading days, the histogram shows estimates significantly different from zero at 5% α-level. "Low frequency" limit orders are defined as those submitted in the prevailing 3 minutes and not cancelled.

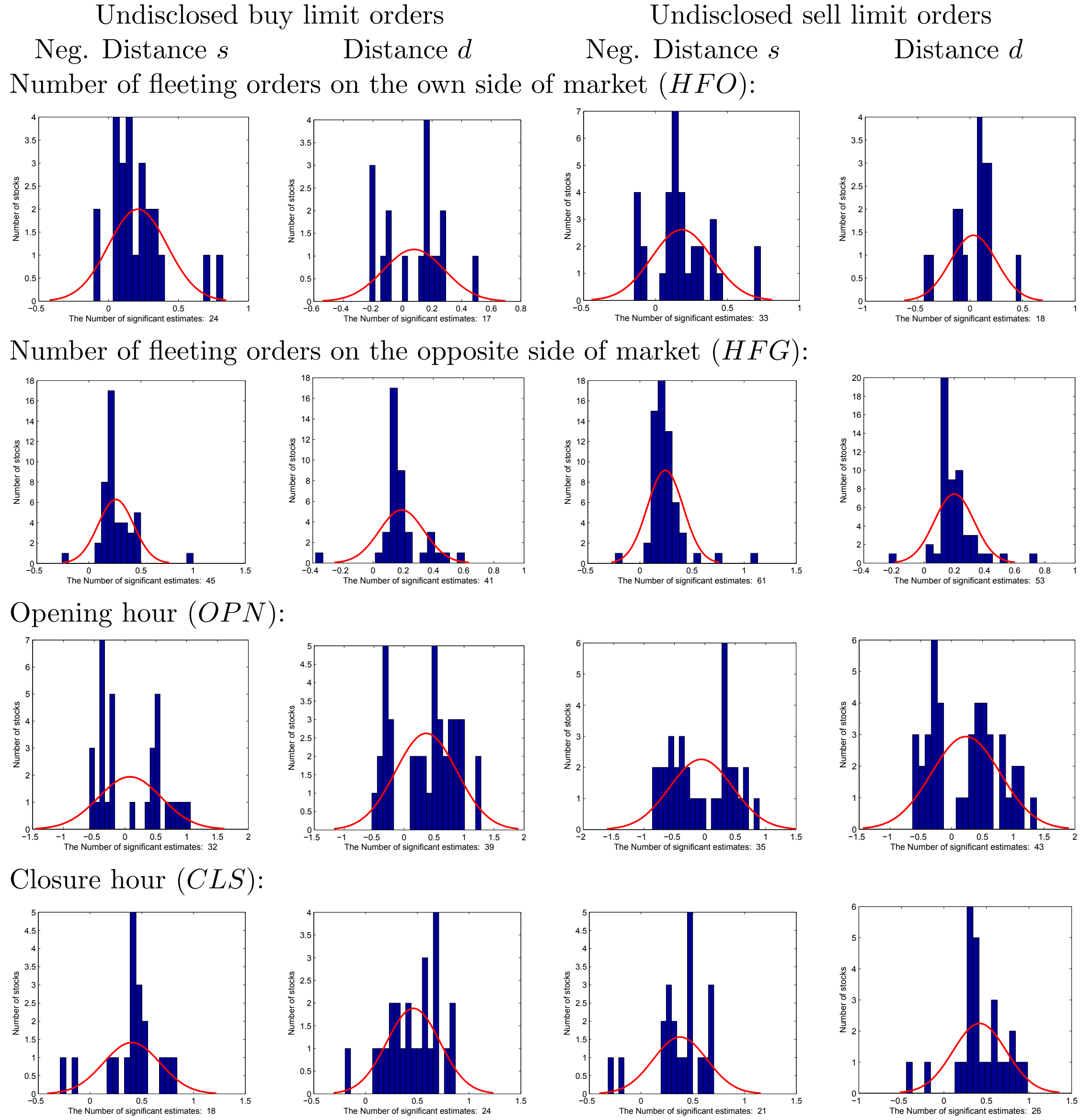


Figure B.4: The histogram of significant estimates. Ordered probit estimates of hidden order locations on the bid and ask side depending on categorized measures s and d as discussed in Section 3.3.4. The order aggressiveness is declining with the category label, thus negative coefficients are associated with increasing aggressiveness. Based on 99 NASDAQ stocks during the period of October 2010 with 21 trading days, the histogram shows estimates significantly different from zero at 5% α-level. The fleeting order is defined as the limit order cancelled in one second after the submission.

Selbständigkeitserklärung

Ich bezeuge durch meine Unterschrift, dass meine Angaben über die bei der Abfassung meiner Dissertation benutzten Hilfsmittel, über die mir zuteil gewordene Hilfe sowie über frühere Begutachtungen meiner Dissertation in jeder Hinsicht der Wahrheit entsprechen.

Berlin, 20 Dezember 2011

Ruihong Huang